TABLE 6

Gross lesions in animals sacrificed 45 days after infection with a 0.2 ml inoculum of *M. bovis* ATCC35721 containing 7.6×10^5 CFU.

Guinea pigs	Injection site	Prefemoral lymph nodes	Spleen
A	+	+	-
B	+	+	-
C	+	+	-

TABLE 7

Gross lesions in animals sacrificed 45 days after infection with a 0.2 ml inoculum of *M. bovis* WAg300 containing 2.8×10^5 CFU.

Guinea pigs	Injection site	Prefemoral lymph nodes	Spleen
A	+	+	+
B	+	+	+
C	+	+	+

M. bovis strains isolated from these animals were shown to be identical to *M. bovis* WAg300 by junction fragment analysis.

The difference between the two sets of guinea pigs with respect to the presence or absence of spleen lesions clearly indicated that *M. bovis* WAg300 was more virulent than *M. bovis* ATCC35721.

F. Isolation of part of the integrated virulence determining cosmid

Genomic DNA was prepared from *M. bovis* WAg300, digested with the restriction enzyme NotI and ligated under conditions favoring self ligation. The ligation

mixture was electroporated into *E. coli*, and kanamycin resistant clones were isolated. A plasmid isolated from one of these clones was denoted pMUA2. This plasmid contained the pYUB178 kanamycin resistance gene and *E. coli* origin of replication from the integrated cosmid in *M. bovis* WAg300 as well as approximately 6 kb of cosmid insert DNA. The relationship between pUHA2 and the original cosmid, designated pUHA1, which was integrated in *M. bovis* WAg300 and which was never isolated in total is shown in Fig. 1.

G. Selection of cosmids with possible virulence determining factors

A 2 kb MluI fragment from the insert of pUHA2 was used as a colony hybridization probe of the *E. coli* pYUB178::*M. bovis* WAg200 library. Approximately one colony in every 130 library colonies gave a positive hybridization signal. Cosmids were isolated from 48 hybridizing clones using standard plasmid preparation methods and compared to each other and to pUHA2 on the basis of restriction enzyme digestion patterns. Three cosmids, designated pUHA3, pUHA4 and pUHA5, had most similarity to pUHA2 and are shown in Fig. 2. Two other cosmids with inserts which overlapped those of pUHA3-pUHA5 were also selected from the remaining 45 cosmids by using pUHA2 as a probe of Southern blots of cosmid restriction digests. These cosmids, designated pUHA6 and pUHA7 are also shown in Fig. 2.

H. Preparation of putative virulence sequences for guinea pig reinoculation

Cosmids pUHA3-pUHA7 were electroporated into *M. bovis* ATCC35721 and clones of *M. bovis* ATCC35721(pUHA3-pUHA7) were recovered using kanamycin selection. These recombinant *M. bovis* clones, designated WAg301-WAg311 were inoculated into guinea pigs to assess

their virulence. The number of *M. bovis* clones inoculated was greater than the number of cosmids because in some cases, junction fragment analysis of individual clones revealed three different patterns were obtained for some cosmids. In cases where more than one pattern was obtained for DNA isolated from clones containing a particular cosmid, subcultures of clones representing each pattern were combined for inoculation. The association between cosmids and *M. bovis* recombinants is shown in Table 1. Guinea pigs that had received *M. bovis* recombinants containing cosmids pUHA3, pUHA4, pUHA5, and pUHA7 developed extensive lung or spleen lesions, indicating that these cosmids had restored the virulence to the *M. bovis* ATCC35721 strain. These three cosmids contain genomic inserts of approximately 40-43 kb and have a common overlapping segment of approximately 10 kb.

Cosmid pUHA3 was partially digested by *Sau3AI* and in separate experiments 2-4 kb and 10-15 kb fragments were cloned into the cosmid shuttle vector pUHA8. Vector pUHA8 was produced from pYUB178 by incorporating *PacI* sites on either side of the *BclI* cloning site. These libraries of pUHA3 were electroporated into *M. bovis* ATCC35721 to produced libraries of *M. bovis* ATCC35721(pUHA8::pUHA3). Approximately 300 colonies from the 2-4 kb library and 1000 colonies from the 10-15 kb library were pooled separately, subcultured and inoculated into guinea pigs.

Guinea pigs that had received *M. bovis* recombinants containing either the 2-4 kb fragments or the 10-15 kb fragments, developed extensive spleen lesions indicating that these fragments had restored virulence to the *M. bovis* ATCC35721 strain. *M. bovis* organisms were isolated from the spleen lesions and subcultured for DNA extraction. DNA prepared from these cultures was digested with *PacI* and electrophoresed on

agarose gels. No restriction fragments could be clearly visualized by staining with ethidium bromide so the gels were Southern blotted onto nylon and hybridized with a DNA probe of the entire insert of pUHA2. This probe revealed two hybridized bands for many of these isolates. One of the bands was the same for all isolates and corresponded to the position on the blot of undigested genomic DNA. The other band varied in size from one isolate to another but in no case was smaller than approximately 3 kb. One strain containing an approximately 3 kb fragment was designated WAg320 and used for further analysis. These results showed that a DNA fragment of approximately 3 kb was sufficient to restore virulence to *M. bovis* ATCC35721. This 3 kb sequence has sufficient overlap with the insert of pUHA2 for detectable hybridization to occur between them. This alignment of the 3 kb sequence and pUHA2 is also consistent with the virulence restoring abilities of cosmids puHA4, pUHA5 and pUHA7 since most of the insert of pUHA2 is within the shared DNA segment of cosmids pUHA4, pUHA5, and pUHA7.

I. Restriction mapping of pUHA3 cosmid

A restriction map of cosmid pUHA3 (Fig. 3) was constructed for the enzymes *Mlu*I, *Nhe*I and *Not*I using a partial digestion technique. The cosmid insert contained no sites for the enzyme *Xba*I, whereas the pYUB178 vector contained two sites as shown (Fig. 3). In the technique used, cosmid pUHA3 was partially digested with each of the three enzymes separately and then the partial digests were digested with *Xba*I. DNA fragments in each partial digest were separated in duplicate by agarose electrophoresis and transferred to nylon filters by Southern blotting. One of the duplicates was hybridized with a ^{32}P labelled probe of the left hand vector arm of

pUHA3 and the other duplicate was hybridized with a probe of the right hand vector arm of pUHA3. Best estimates of the molecular size differences between the labelled fragments were obtained by comparison to labelled DNA markers and these were also compared to fragment sizes of complete digests of pUHA3 with the same enzyme.

J. Sequencing of 3 kb sequence

WAg320 was digested with PacI and the 3 kb fragment was ligated into the PacI site of the sequencing vector pUHA9 using standard methods. The "Erase-a-base" system (Promega) was used to make progressive, unidirectional deletion mutants of two clones designated pUHA11 and pUHA16 which contained the 3 kb fragment in opposite orientations. Appropriately sized deletion mutants were cloned and chosen as instructed by the manufacturer's protocols. Polymerase chain reaction sequencing was performed by using commercial kits (Gibco-BRL and Intermed) in accordance with the manufacturer's instructions. The 2745 bp fragment that restores virulence to *M. bovis* ATCC35721 is shown in Figure 9. Figure 9A shows this sequence together with a 530 amino acid translation of the largest ORF. The first codon of this ORF at positions 835-837 is contiguous with the likely ribosome binding site so initiation may actually occur at codon three at positions 841-843.

K. Comparison of the 3 kb Mycobacterial DNA sequence with GenBank sequences

The DNA sequence obtained from the 3 kb fragment that restores virulence to *M. bovis* ATCC35721, shown in Figure 9, was analyzed using the 7.3.1-UNIX update (September 1993) of the program package supplied by the University of Wisconsin Genetics Computer Group (575 Science Drive, Madison, Wisconsin 53711); this

package is abbreviated as "GCG". An earlier version of the package is described in Devereux, J., et al., (1984), Nucl. Acids Res. 12: 387-395.

The comparison was performed as follows. The DNA sequences of the contigs were translated into amino acids (using the program TRANSLATE) and compared to the GenBank database update 82.0 using the programme TFASTA. This comparison revealed that the sequence analyzed had significant homology with numerous sigma factors. Some of the DNA sequences of the sigma factors with which the homology was particularly high were obtained from the GenBank database using the programme FETCH and their coding sequences were translated into amino acids using TRANSLATE. These sigma factors were then compared to an amino acid translation (using TRANSLATE) of the large ORF on the largest contig using the programme PILEUP. A smaller downstream contig was also translated using TRANSLATE and compared in the same PILEUP comparison. FETCH, PILEUP, TFASTA and TRANSLATE are programmes in the GCG package.

The results of a PileUp comparison of hrdB principal sigma factors from *Streptomyces coelicolor* (GenBank Accession No. X52983) and *Streptomyces griseus* (GenBank accession No. L08071) with the amino acid translation of the ORF from the *M. bovis* virulence restoring factor is shown in Figure 10-A. It can be seen from the results that there is a high degree of relatedness between all three sequences, particularly in the region above 290.

Figure 11 presents the results of a GAP comparison of *Streptomyces griseus* principal sigma factor (Peptide translation of GenBank accession No. L08071 from nucleotide numbers 570 to 1907, which is the coding sequence of the *hrdB* gene) with peptide translation of the large ORF of the approximately 3 kb DNA fragment from

M. bovis associated with virulence. Exact homology between the sequences is indicated by vertical dashes.

While there were significant homologies of the sequences encoded in the *M. bovis* fragment with the sigma factor sequences indicated above, the overall homology detected was less than about 65% to 70% with any specific sequence. In addition, there was no exact match with any of the GenBank sequences.

L. Identification of a Mutation Associated with Avirulence

The 2.7 kb fragment from *M. bovis* WAg200 was sequenced on both chains using an ordered deletion mutant strategy and polymerase chain reaction sequencing with ^{33}P. A probe of this fragment was used to select hybridizing clones from replica plates of genomic libraries of *M. bovis* ATCC35721, *M. bovis* WAg201 (another virulent New Zealand strain), and *M. tuberculosis* Erdman. The homologous DNA fragments were isolated and sequenced and their large ORFs translated for the PILEUP comparison.

The sequence of the 2.7 kb fragment encoding the *rpoV* gene from *M. bovis* WAg200 and comparison of its translation with those of other *M. bovis* and *M. tuberculosis* *rpoV* genes and principal sigma factors from two *Streptomyces* species is shown in Figure 12. Figure 12a presents the sequence of *M. bovis* WAg200 showing the large ORF which begins with GTG at position 835-837. Since the potential ribosome binding sites (underlined) are so close or overlap this codon, the likely initiation site is the third codon of the ORF, as indicated. The three mutations in *M. bovis* ATCC35721 and their effect on the translation of *rpoV* are shown respectively above and below the equivalent sequences from *M. bovis* WAg200. Two of the three mutations are also found in one or more of

the other *M. tuberculosis* complex strains analyzed (strain numbers in brackets).

Figure 12b presents a comparison of putative principal sigma factors of four *M. tuberculosis* complex strains and two *Streptomyces sp.* Upper case letters denote amino acids that agree with the consensus sequence of the *M. tuberculosis* complex. An arrow denotes the position of the amino acid in the *M. bovis* ATCC35721 sequence that differs from that of all three of the other *M. tuberculosis* complex strains. These results indicate that it is this difference that causes *M. bovis* ATCC35721 to become avirulent. This position is highly conserved among principal sigma factors and their homologues and the region in which it occurs has the characteristics of a helix-turn-helix motif and is believed to be involved in -35 sequence recognition. (Lonetto, M. et al. (1992), J. Bact. 174:3843-3849). Mutation of an arginine to a histidine in this region has previously been shown to cause an alteration in promoter recognition in *Eschicherichia coli* (Gardella, T., et al. (1989), J. Mol. Biol. 206:579-590). But mutation at the equivalent position in the *M. bovis* ATCC 35721 sequence has not been reported.

Example 2

POLYNUCLEOTIDES ENCODING VIRULENCE FACTORS ISOLATED BY A MOUSE COMPLEMENTATION ASSAY

A method for identifying virulence determinants by genetic complementation was discovered that requires: (i) two strains that are genetically similar; (ii) a phenotype associated with virulence; and (iii) gene transfer systems. An existing pair of *M. tuberculosis* strains, H37Rv (virulent) and H37Ra (avirulent), distinguishable by their ability to cause disease in animal models were used. H37Ra and H37Rv were derived

from the same clinical isolate in 1934 and pulsed field gel analyses of DNA fragments generated by digestion with infrequently cutting enzymes revealed that their macroscopic genome organization was similar. The well-characterized difference in growth rates in mouse lungs and spleens of H37Ra and H37Rv correlated with their pathogenicity. The ability of H37Ra/H37Rv recombinants to grow faster than H37Ra in the mouse was defined as a potential virulence phenotype.

A genomic library of *M. tuberculosis* H37Rv was constructed in an integrating cosmid vector, pYUB178, and electroporated into H37Ra. Mice were infected with pools of H37Ra recombinants containing H37Rv DNA to allow the selection of growing clones in mouse spleen and lung. The integrating shuttle cosmid libraries, based on the mycobacteriophage L5 integration system, were ideal for *in vivo* complementation because: (i) only approximately 225 clones were required to represent the H37Rv genome, (ii) toxic effects associated with the expression of genes from multicopy plasmids were avoided, (iii) kanamycin selection pressure was not necessary to maintain the cosmid, and (iv) clusters of contiguous genes can be delivered and expressed.

The growth rates of selected recombinants were measured in mouse spleen and lung, and a method was developed to retrieve the H37Rv insert DNA from the chromosome of a recombinant. This method allowed for the identification and characterization of a 25 kb DNA fragment of *M. tuberculosis* which conferred an *in vivo* growth advantage to the growth-defective H37Ra.

A. Bacterial strains and growth conditions

M. tuberculosis strains H37Ra and H37Rv were provided by Wilbur Jones of the Centers for Disease Control, Atlanta, and were grown in enriched 7H9 broth

[Middlebrook 7H9 medium enriched with albumin-dextrose complex (ADC) or oleic acid-albumin-dextrose complex (OADC) (Difco Laboratories, Detroit, Mich.) and a 0.05% polyoxyethylene sorbitan monooleate (Tween-80®)], under Biosafety Level 3 (BSL3) containment. All cultures were grown at 37°C. *E. coli* strains χ2764 (13), HB101 (4)and DH5α (Bethesda Research Laboratories Life Technologies Inc., Gaithersburg, MD) were grown in L broth. Strain χ2764 was grown at 30°C. See Table 8 for a list of strains and plasmids.

B. Construction of shuttle cosmid and H37Rv library

The pYUB178 integrating shuttle cosmid (Figure 1A), was constructed by ligating the 975 bp cos-containing *BglII/BclI* fragment of lambda DNA to the *BclI*-digested, calf-intestine alkaline phosphatase (CIP)-treated (Boehringer Mannheim Biochemicals, Indianapolis, IN) pMV305.F (18, 27) under conditions which favored the formation of linear concatemers, i.e. greater than 50 ng/μl final DNA concentration.

Genomic DNA of H37Rv was prepared by mechanical disruption of bacterial cells and subsequent phenol-chloroform extractions as previously described (12). H37Rv genomic DNA was partially digested with a range of concentrations of *Sau3AI* to generate 30-50 kb-sized fragments. Fragments of 30-50 kb were isolated as previously described (14). The 30-50 kb *Sau3AI* fragments of chromosomal DNA were then ligated to CIP-treated, *BclI*-digested pYUB178 DNA; the final DNA concentration was 50-100 ng/μl and the DNA molar ratio of insert to vector was 1.

C. Library packaging into lambda phage heads and tails

Four μl of a ten μl ligation mixture was in vitro-packaged with the GigaPack II Packaging Extract

(Stratagene, La Jolla, CA) according to the manufacturer's procedure. The *in vitro*-packaged lysate was transduced, using previously described methods (14), into the *in vivo* packaging strain of *E. coli* χ2764 (13).

5

D. *In vivo*-packaging

The 10^3-10^4 kanamycin-resistant recombinant clones were pooled and inoculated into L broth containing 25 µg/ml kanamycin. One aliquot was grown to prepare plasmid DNA by an alkaline lysis method. The other aliquot was grown by *in vivo*-packaging which was accomplished by previously described procedures (13). The titer of the lysate prepared from χ2764 transductants containing the pYUB178::H37Rv library was approximately 1×10^9 cfu/ml. The lysate was stored at 4°C after filtering through a 0.45 µm pore sterile filter.

E. Construction of H37Ra (pYUB178::H37Rv) recombinant pools.

An eight day old H37Ra culture was electroporated with the pYUB178::H37Rv library DNA in plasmid form, and separately, with pYUB178 DNA. Approximately 450 transformants arose from five independent electroporations of cells with approximately 1 µg library DNA each. Two pools of H37Ra (pYUB178::H37Rv) recombinants, pool 1 and pool 2, were made by collecting and inoculating approximately 225 colonies into 50 ml of enriched 7H9 broth containing 10 µg/ml kanamycin, and allowing growth for approximately two weeks. Aliquots of pools were distributed and frozen in cryovials for later use in animal experiments.

Another pool of H37Ra(pYUB178::H37Rv) recombinants, pool 3, consisted of approximately 260 clones and was used to determine whether the pools were representative. Recombinants of pool 3 were collected directly from plates of enriched Middlebrook 7H10 agar

containing 25 μg/ml kanamycin after growth following electroporation; an aliquot was inoculated into enriched 7H9 broth without kanamycin and allowed to grow standing at 37°C for approximately two weeks. Total DNA was isolated from pool 3 before and after growth in broth. DNA was subjected to Southern analysis using the 1.1. kb DraI/SspI DNA fragment of pYUB178 as a probe.

F. Mouse infection

In experiments J2, J2P, J5 and J5P that used the mouse to select individual recombinant clones from pools 1 and 2, and in subsequent growth measurement experiments, J33 and J36, groups of C57BL/6 mice aged 6-8 weeks were intravenously inoculated with 0.2 ml of each culture tested. Five mice were inoculated with each recombinant group or control group per timepoint. Inoculation of mice with spleen-passaged bacteria was accomplished by first homogenizing spleens after fourteen days infection in 5 ml sterile saline. One ml of the 5 ml spleen homogenate from each of the five mice per group was pooled and filtered through sterile gauze to exclude tissue clumps. The filtrate was used to directly inoculate another set of mice in experiments J2P and J5P. See Table 9 for details or mouse experiments.

Individual colonies that grew from plated lung homogenates in experiments J2P and J5P were picked and grown in enriched 7H9 broth for subsequent mouse experiments and DNA analyses.

G. Retrieval of pYUB178::H37Rv cosmids from chromosomes of in vivo-selected recombinants

Chromosomal DNA was isolated from individual H37Ra (pYUB178::H37Rv) recombinant clones using chemical disruption of bacterial cells as previously described (28). DNA was partially digested with Sau3AI; fragments of 30-50 kb were size-fractionated and eluted from

agarose gels as described above. The 30-50 kb fragments were ligated to the 975 bp *BglII/BclI* fragment containing *cos* of coliphage lambda DNA. The ligation conditions were such that the final DNA concentration was 50 to 100 ng/μl, and the molar ratio of chromosomal DNA fragments to *cos* DNA fragments was 1.

The ligation mixture was packaged into lambda phage heads and tails using the Stratagene GigaPack kit, and transduced into *E. coli* strain NB101. Individual kanamycin-resistant transductant colonies were picked and cosmid DNA was isolated. Cosmid DNA was then analyzed by restriction digestion and Southern hybridization.

K. Restriction and Southern analyses of retrieved cosmids

Digested cosmid DNA was subjected to agarose gel electrophoresis in 0.8% agarose in TAE buffer. DNA was Southern blotted from gels onto nylon membranes by capillary diffusion, UV-crosslinked and hybridized with probes derived from pYUB178. Probes consisted of either the 1.1 kb *DraI/SapI* fragment of pYUB178, or the 436 bp *AseI/BclI* fragment of pYUB178 that contained lambda DNA adjacent to *cos*, or the 756 bp *AseI/BclI* fragment of pYUB178 that contained part of *aph*. Probes were labeled with {α-^{32}P}dCTP using random hexamer priming with the Pharmacia oligolabeling kit (Pharmacia LKB Biotechnology AB, Uppsala, Sweden), or with horseradish peroxidase according to the protocol of the Enhanced Chemiluminescence ECL Gene Detection System (Amersham International, Amersham, UK).

L. Screening the pYUB178::H37Rv library in *E. coli*

The pYUB178::H37Rv library DNA lysate, 10^9 cfu/ml, was serially diluted to a concentration of 10^4 cfu/ml in SM buffer {50 mM Tris-Cl (pH 7.5), 100 mM NaCl, 8 mM $MgSO_4 \cdot 7H_2O$}, and transduced into *E. coli* strain

HB101. Aliquots of infected cells were plated onto L agar containing 25 μg/ml kanamycin such that each plate would contain approximately 150 colonies. After overnight incubation at 37°C, colonies from each plate were lifted onto Biotrans nylon filters (ICN Biomedicals, Inc., Irvine, CA). The filters were treated with denaturing buffer and neutralization buffer and UV-crosslinked. A probe was made from a cosmid, pYUB352, derived from the mc^2806 recombinant clone. The cosmid pYUB352 was linearized by digestion with *Ase*I and labeled with [α-^{32}P]dCTP. Filters were hybridized overnight according to the manufacturer's protocol (ICN Biomedicals, Inc.).

Thirty hybridizing clones were picked and streaked onto plates, and subjected to secondary screening with the pYUB352 probe. Ten strongly hybridizing clones were picked and analyzed by Southern hybridization with pYUB352 as a probe. Four cosmids, two that shared H37Rv restriction fragments with pYUB352, and two that did not share H37Rv restriction fragments with pYUB352, were electroporated individually into H37Ra.

I. In vivo growth of pYUB352-overlapping and -nonoverlapping recombinants

Single H37Ra transformant colonies from each of the four electroporations were grown in enriched 7H9 broth containing kanamycin to prepare sufficient culture for mouse experiments. The *in vivo* growth rates of H37Ra containing pYUB352-overlapping and -nonoverlapping clones were measured in the experiment designated J36 (see Table 9).

K. Results

i. Construction of shuttle cosmid and H37Rv library

The integrating cosmid pYUB178 contains an *E. coli ori* derived from pUC19, the L5 attP site, the L5 integrase gene, a kanamycin resistance gene, *aph*, derived from Tn*903*, the lambda *cos* sequence and a unique cloning site, *BclI* (see Figure 4A). The L5 mycobacteriophage attachment site attP, and integrase gene, *int*, mediate site-specific integration into the mycobacterial chromosome (18). The H37Rv library was constructed by ligating 40 kb size-selected chromosomal DNA fragments, generated by partial digestion with *Sau3AI*, to alkaline phosphatase-treated pYUB178, linearized by digestion with *BclI*. The ligation mix was packaged into lambda phage heads and tails, and transduced into *E. coli*. The approximately 4000 kanamycin-resistant transductant colonies were theoretically enough to represent the H37Rv genome forty times. Twelve individual cosmids of the H37Rv library were isolated from randomly picked *E. coli* transductant colonies and examined by restriction analyses. No two cosmids were alike, and each cosmid had an insert size of 35-40 kb (data not shown). The H37Rv library DNA was isolated as plasmid from the complete pool of *E. coli* transductants and electroporated into H37Ra. To identify the H37Rv insert within the chromosome of a H37Ra(pYUB178::H37Rv) recombinant, a method to detect the H37Rv DNA fragments immediately adjacent to pYUB178 sequences was devised. The method of analysis depicted in Figure 4B allows the identification of *PstI* restriction fragments of the H37Rv DNA at the junctions of pYUB178 sequences on either side of the *BclI* cloning site (see Figure 4B). The pYUB178-H37Rv junctional fragments of individual H37Ra(pYUB178::H37Rv)

recombinants are visible as bands in the Southern analysis in Figure 4C, lanes 1-3.

To determine if a representative panel of H37Ra(pYUB178::H37Rv) recombinants was generated, approximately 260 transformant colonies, pool 3, were collected after growth on kanamycin-containing 7H10 agar; an aliquot of pool 3 was transferred to enriched 7H9 medium and allowed to grow for approximately two weeks. Chromosomal DNA was isolated from pool 3 both before and after growth in broth. These DNAs were subjected to *PstI* digestion and agarose gel electrophoresis, followed by transfer to a nylon membrane and hybridization to a pYUB178 probe (Figure 4C). In figure 4C, the smears in lanes 4 and 5 reveal that the pool of H37Ra(pYUB178::H37Rv) recombinants consisted of members having different H37Rv DNA inserts, both before and after growth in broth, suggesting that the pools were representative and stable in the absence of kanamycin selection pressure.

ii. Enrichment and selection of putatively virulent recombinants from pools

Mice were intravenously infected with either H37Ra(pYUB178::H37Rv) recombinant pool 1 or 2. Two weeks post-infection, mouse spleens were individually homogenized, pooled, and used to infect a second group of mice. Individual recombinant colonies that grew from the plated lung homogenates prepared from the second group of mice were picked. To characterize the integrated cosmid in each recombinant, chromosomal DNAs were isolated from these individual recombinants and subjected to Southern analysis with a pYUB178 probe. The junctional fragment analyses of selected individual recombinants from the in vivo-passed pool 2 in experiment JSP (see Table 9) are shown in Figure 4C, lanes 1, 2 and 3. Lane 1 shows the clone designated mc^2807, lane 2 shows the clone

designated mc^{2}806, and lane 3 shows a clone that has junctional fragments identical to those of mc^{2}806. Because clones having junctional fragments identical to those of mc^{2}806 were isolated from many animals during two different experiments, J2P and J5P, (data not shown), mc^{2}806 was further characterized.

iii. *In vivo* growth rate comparisons

Growth rate comparisons of clones mc^{2}806, mc^{2}816 (H37Ra containing pYUB178, see Table 9) and H37Rv were made (see Figure 5). Clone mc^{2}806 grew in the spleen at a rate that was slightly lower than the growth rate of H37Rv during the first two weeks of infection. Clone mc^{2}816 barely grew. After the initial growth phase, mc^{2}806 was cleared from the spleen at a lower rate than the rate of clearance of mc^{2}816. H37Rv persisted at its day 28 level, at least through the experimental endpoint, day 84. Clone mc^{2}806 did not grow faster than mc^{2}816 during the first two weeks in mouse lung (Figure 5B). Therefore the faster *in vivo* growth rate of mc^{2}806 compared to mc^{2}816 was evident only in mouse spleen. The growth rates of mc^{2}806, mc^{2}816, and H37Rv in enriched 7H9 broth were virtually identical (data not shown).

iv. Identification of a H37Rv DNA insert that confers a faster *in vivo* growth rate to H37Ra

To prove that the H37Rv DNA insert present in an *in vivo*-selected recombinant was responsible for its *in vivo* growth phenotype, it had to be retrieved from the chromosome. A disadvantage of the stably integrating pYUB178::H37Rv cosmid library is the difficulty of cosmid retrieval from the chromosome of a H37Ra(pYUB178::H37Rv) recombinant; the excision functions of L5 are not yet understood. Hence, a method was devised to clone the H37Rv DNA insert as a cosmid (see Figure 6A). The lambda *in vitro*-packaged ligation mix that contained random

pieces of the mc^2806 chromosome was transduced into *E. coli* for the purpose of selecting H37Rv DNA-containing cosmids. Only those cosmids containing the *E. coli* and *aph* replicated under kanamycin selection pressure (cf Figure 6A). The Southern analyses of 16 of the 33 retrieved cosmids of mc^2806 from *E. coli* transductants is shown in Figure 6B. The cosmids were digested with both *EcoRI* and *AseI* and analyzed by gel electrophoresis. The 434 bp probe, generated by digestion of pYUB178 with *AseI* and *BclI*, hybridized to the H37Rv/pYUB178 junction that included lambda DNA adjacent to *cos*. By comparing the sizes of the junctional fragments of the retrieved cosmids with the sizes of the junctional fragments of mc^2806 in lane 1, one can determine whether the entire H37Rv insert DNA has been retrieved. Only one of the 16 cosmids did not contain the full-sized H37Rv fragment adjacent to the pYUB178 junction (Figure 6B, lane 14). The retrieval procedure was very efficient; 32 of the 33 mc^2806-retrieved cosmids contained the entire H37Rv insert (data not shown). The cosmid clone designated pYUB352 in lane 15 was used for further study.

v. Identification of pYUB352-overlapping cosmids from the pYUB178::H37Rv DNA library

To prove that the H37Rv insert DNA was responsible for the spleen growth phenotype, it had to be reintroduced into H37Ra and tested. Reintroduction of the H37Rv insert DNA from the mc^2806 recombinant into H37Ra required a replicating vector. Retrieved cosmids did not have the ability to replicate in mycobacteria because they lost the *int* gene when they were removed from the chromosomes of the recombinants. Therefore, pYUB352 DNA was used as a probe to screen the pYUB178::H37RV library in *E. coli* for the H37Rv DNA insert associated with mc^2806. Colonies of *E. coli* (pYUB178::H37Rv) library transductants were transferred

to nylon filters, lysed, and probed with pYUB352 DNA. Cosmids that shared H37Rv DNA with pYUB352, designated pYUB353 and pYUB354, and unrelated cosmids, designated pYUB355 and pYUB356, were separately transformed into H37Ra.

vi. The H37Rv DNA of mc^2806 confers *in vivo* growth advantage to H37Ra

The growth rates of H37Ra recombinant clones containing pYUB352-overlapping and -nonoverlapping cosmids were tested in mice (experiment J36, see Table 9). The H37Ra recombinants containing the pYUB352-overlapping cosmids grew as well as mc^2806, and the H37Ra recombinants containing pYUB352-nonoverlapping cosmids grew poorly or did not grow at all (Figure 7). These data indicate that the H37Rv DNA that is shared by pYUB352, pYUB353, and pYUB354 expresses a gene or gene(s) associated with growth in the spleen.

vii. Mapping the *ivg* region of H37Rv

The pYUB352, pYUB353, and pYUB354 cosmids were mapped by restriction digest and analyzed by Southern hybridization (see Figure 8). The schematic of Figure 8C shows the physical map of the H37Rv DNA insert of each clone. A DNA region of approximately 25 kb is shared between the clones. This region was designated *ivg* or *in vivo* growth advantage.

TABLE 8

Bacterial strain or clone	Description	Source
E. coli		
HB101	F-ara14 *leuB6 proA2 lacY1 glnV44 galK21-recA13 rpsL20 xyl-5 mtl-1 thi-1 hsdS20*	(3)
χ^{2764}	HB101 lysogenized with λ c1857 *b2 red3 S7*	(8)
DH5α	F-endA1 *hsdR17 supE44 thi-1 l-recA1 gyrA96 relA1* Δ *(argF-lacZYA) U169 Φ80dlacZ ΔM15*	BRL, Inc.
M. tuberculosis		
mc^2806	H37Ra containing pYUB178::H37Rv *ivg*	This study
mc^2822	H37Ra containing pYUB353	This study
mc^2823	H37Ra containing pYUB354	This study
mc^2824	H37Ra containing pYUB355	This study
mc^2825	H37Ra containing pYUB356	This study
Shuttle Plasmid		
pYUB178	Integrating shuttle cosmid vector	This study
pYUB352	H37Rv *ivg*-containing cosmid derived from mc^2806	This study
pYUB353	pYUB178::H37Rv *ivg*	This study
pYUB354	pYUB178::H37Rv *ivg*	This study
pYUB355	pYUB178::H37Rv	This study
pYUB356	pYUB178::H37Rv	This study

TABLE 9

Experiment	Pools and Clones Tested	Inocula (cfu/mouse)	Timepoints (day)
J2	Pool 1 Pool 2 mc^2816	2×10^5 6×10^5 1×10^6	1, 14, 28
J5	Pool 1 Pool 2 mc^2816 H37Rv	1×10^5 6×10^5 1×10^6 6×10^4	1, 14, 28
*J2P	Pool 1 Pool 2 mc^2816	5×10^2 7×10^2 5×10^2	1, 14
*J5P	Pool 1 Pool 2 mc^2816	9×10^2 7×10^2 6×10^3	1, 14
J33	mc^2806, mc^2816, H37Rv	$1\text{-}2 \times 10^4$ 4×10^4 5×10^4	1, 14, 28, 84
J36	mc^2806, mc^2822, mc^2823, mc^2824 mc^2825, mc^2816, H37Rv	1×10^4 $1\text{-}2 \times 10^4$ $1\text{-}3 \times 10^4$ 5×10^4 6×10^4 8×10^4 4×10^4	2, 14, 28, 87

*For J2P and J5P, inocula were estimated from cfu retained in the spleen on day 1; spleen retention is usually 10% of the inoculating dose.

CLAIMS

WE CLAIM:

1. A method for identifying a DNA sequence or sequences associated with virulence determinants in *M. tuberculosis* and *M. bovis* and similar DNA sequences in other tuberculosis complex strains and in strains of other mycobacterial species and in species of other pathogenic organisms comprising the steps of:

a) preparing a genomic DNA library of the pathogenic organism;

b) constructing an integrating shuttle vector containing genomic inserts prepared in step a);

c) transforming via homologous recombination a population of avirulent organisms;

d) isolating the recombinants;

e) inoculating a subject with an adequate inoculum of the recombinants in order to select virulent recombinants;

f) isolating the virulent recombinants; and

g) identifying the DNA insert which confers virulence.

2. A method according to claim 1 wherein the individual inoculated is a mouse.

3. A method according to claim 1 wherein the individual inoculated is a guinea pig.

4. An isolated polynucleotide comprised of a segment of less than 3kb that is essentially homologous to a mycobacterial DNA sequence associated with virulence in mycobacteria, wherein the mycobacterial DNA sequence encodes a sigma factor.

5. An isolated polynucleotide comprised of a segment of less than 3 kb that encodes a polypeptide or fragment thereof, wherein the polypeptide is associated with virulence in mycobacteria and is a sigma factor.

6. An isolated polynucleotide according to claim 5, wherein the polypeptide is essentially homologous to the polypeptide encoded in Figure 9.

7. An isolated polynucleotide comprised of at least 15 sequential nucleotides homologous to a sequence of polynucleotides in Figure 9.

8. A recombinant polynucleotide comprised of a sequence of at least 15 sequential nucleotides homologous to a sequence of polynucleotides in Figure 9.

9. A recombinant polynucleotide comprised of a segment of less than 3 kb that encodes a polypeptide or fragment thereof, wherein the polypeptide is associated with virulence in mycobacteria and is a sigma factor.

10. An expression vector comprised of the recombinant polynucleotide of claim 9.

11. An isolated polynucleotide comprised of a linear segment of at least 15 nucleotides that is substantially homologous to mycobacterial DNA in a plasmid selected from the group consisting of pUHA1, pUHA2, pUHA3, pUHA4, pUHA5, pUHA6, pUHA7, pUHA11, pUHA16, pYUB352, pYUB353, and pYUB354.

12. A host cell comprised of a polynucleotide selected from the group consisting of the polynucleotide

of claim 1, claim 2, claim 3, claim 4, claim 5, claim 6, claim 7, claim 8, and claim 9.

13. A host cell comprised of a polynucleotide according to claim 11.

14. A host cell comprised of the expression vector of claim 10.

15. A diagnostic kit comprised of a polynucleotide and a buffer packaged in suitable vials, wherein the polynucleotide is selected from the polynucleotides according to claims 3, 4, 5, 6, 7, 8, and 9.

16. An isolated polypeptide substantially homologous to a polypeptide associated with virulence in mycobacteria or a fragment thereof, wherein the mycobacterial polypeptide is a sigma factor.

17. The isolated polypeptide of claim 16, wherein the mycobacterial polypeptide is encoded in a DNA sequence shown in Figure 9.

18. An isolated polynucleotide comprised of a segment of less than 3kb that is essentially homologous to a mycobacterial DNA sequence associated with avirulence in mycobacteria, wherein the mycobacterial DNA sequence encodes a sigma factor.

19. A method for producing an altered property in a wild-type bacterial strain other than *M. bovis* comprising mutagenizing a principal sigma factor in the bacteria, wherein the mutagenizing results in converting an arginine to a histidine in the principal sigma factor,

and wherein the conversion occurs at a similar position to that present in *M. bovis* ATCC 35721.

20. The method of claim 19 wherein the mutagenizing results in altered virulence properties of the resulting bacterial strain.

21. A method of using a bacterial strain prepared by the method described in claim 20, the method comprising preparing a vaccine by mixing a pharmacologically effective dose of the strain with a pharmaceutically acceptable suitable excipient.

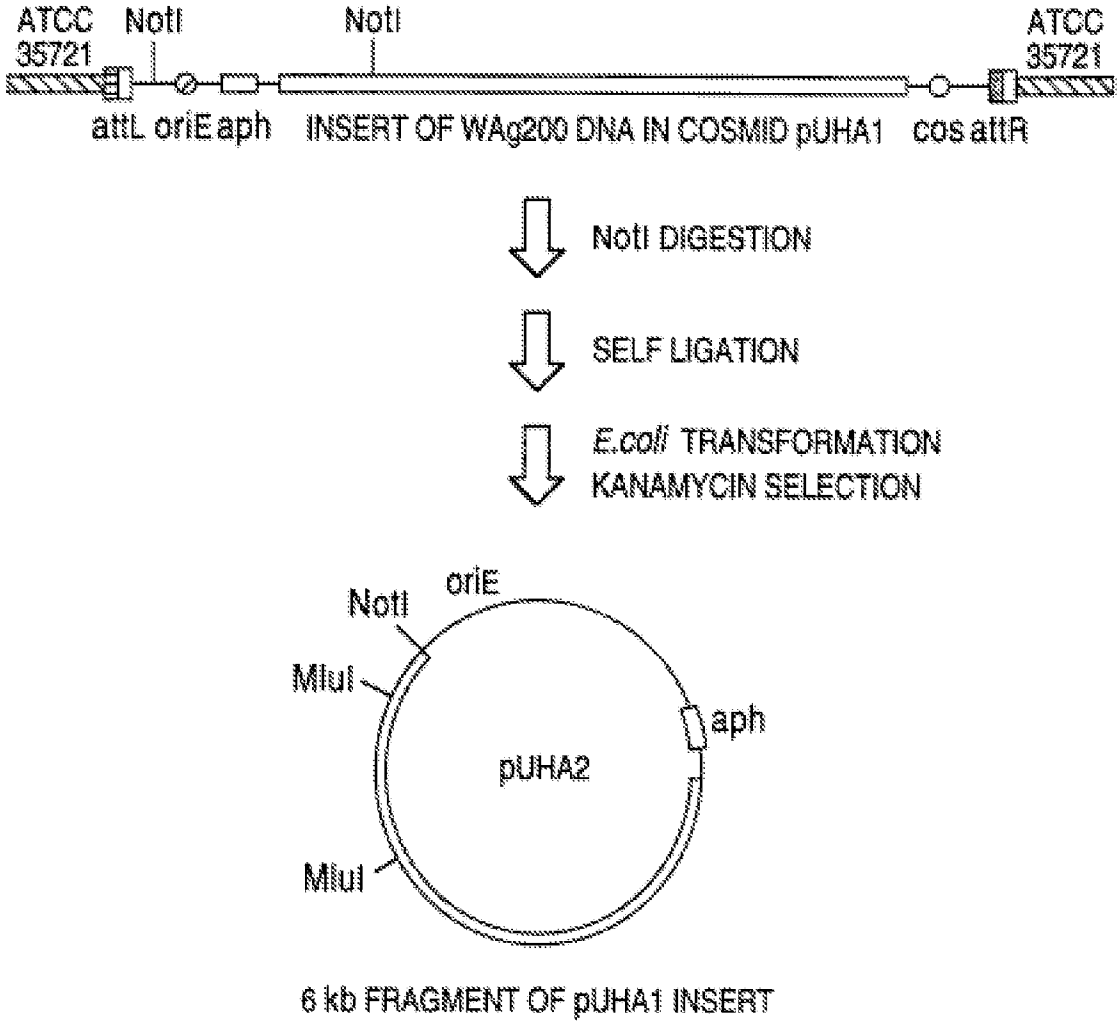
ATCC 35721
NotI
NotI
ATCC 35721
attL oriE aph
INSERT OF WAg200 DNA IN COSMID pUHA1
cos attR
NotI DIGESTION
SELF LIGATION
E.coli TRANSFORMATION
KANAMYCIN SELECTION
NotI
oriE
MluI
aph
pUHA2
MluI
6 kb FRAGMENT OF pUHA1 INSERT

FIG. 1

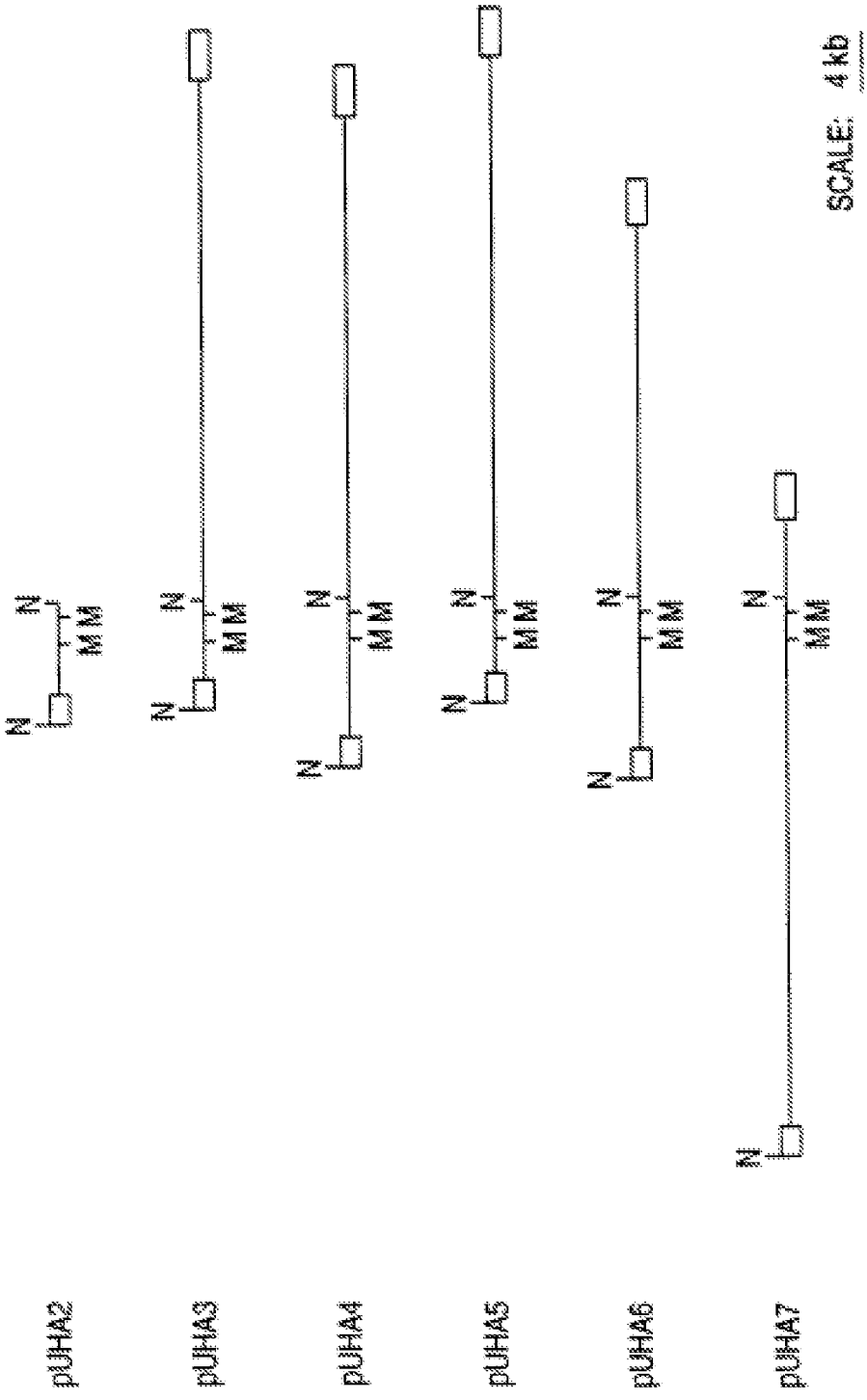

FIG. 2

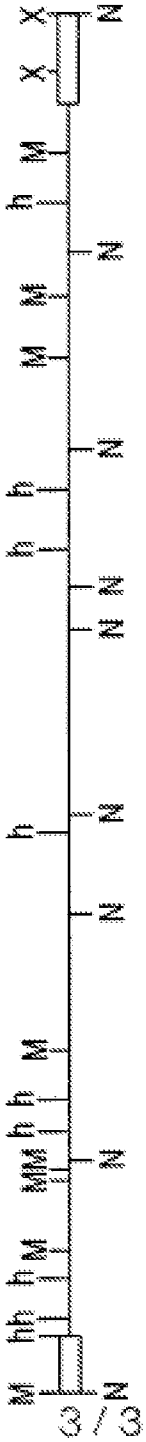

FIG. 3

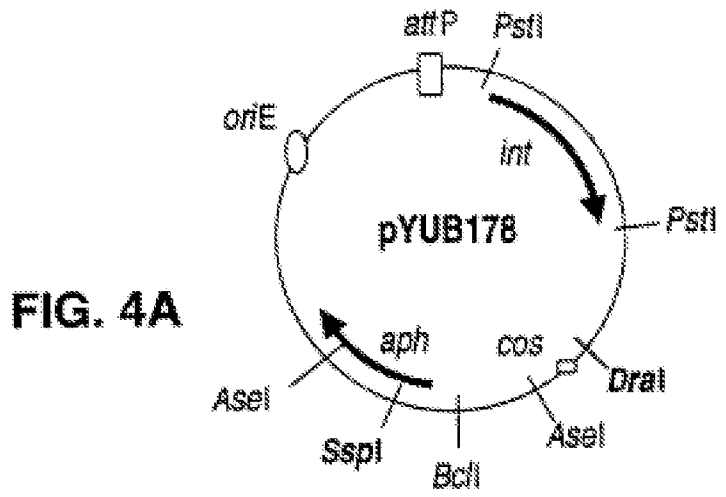
attP
PstI
oriE
int
pYUB178
PstI
aph
cos
DraI
AseI
SspI
BclI
AseI

FIG. 4A

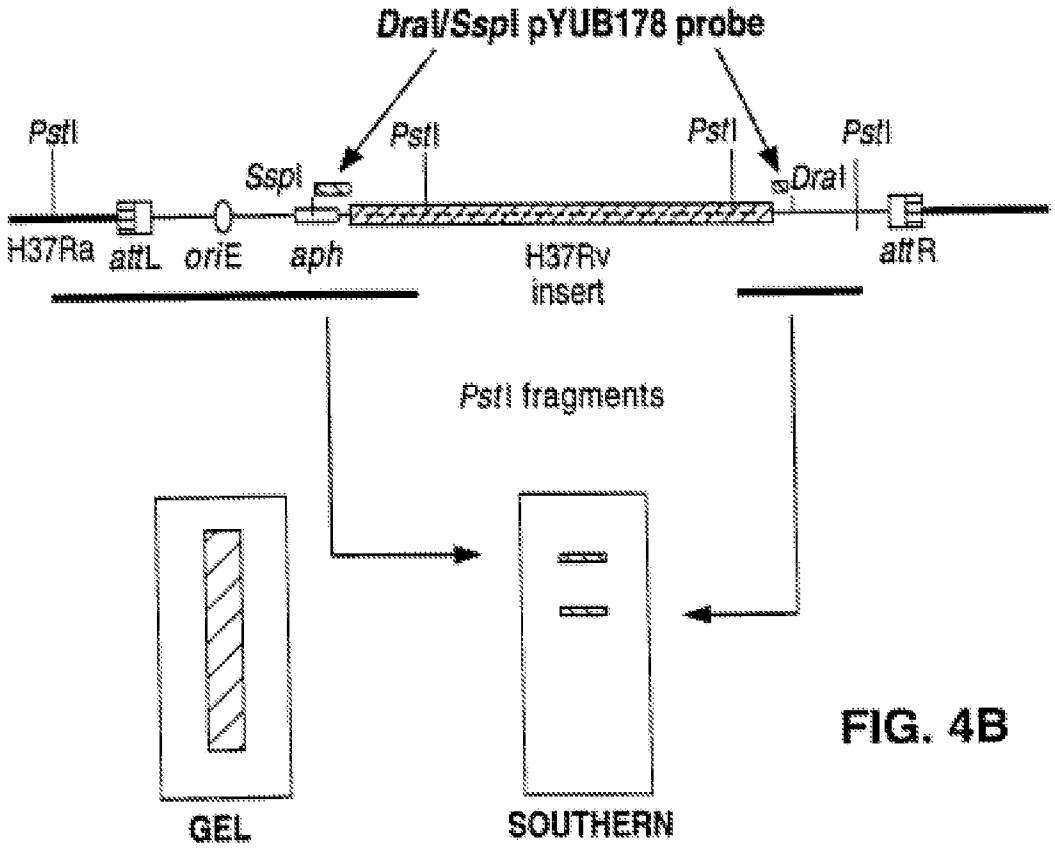
DraI/SspI pYUB178 probe
PstI
PstI
PstI
PstI
SspI
DraI
H37Ra
attL
oriE
aph
H37Rv
insert
attR
PstI fragments
GEL
SOUTHERN

FIG. 4B

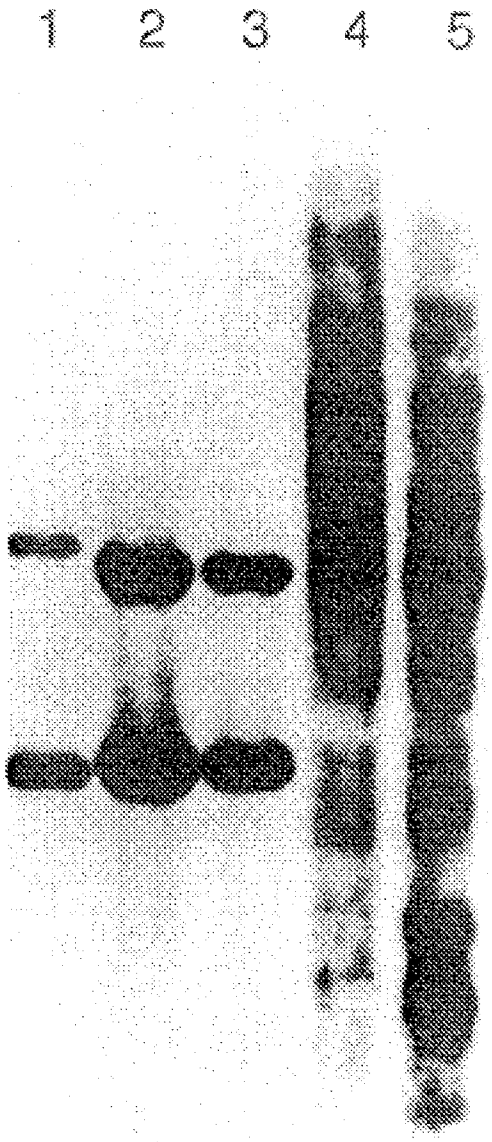

FIG. 4C

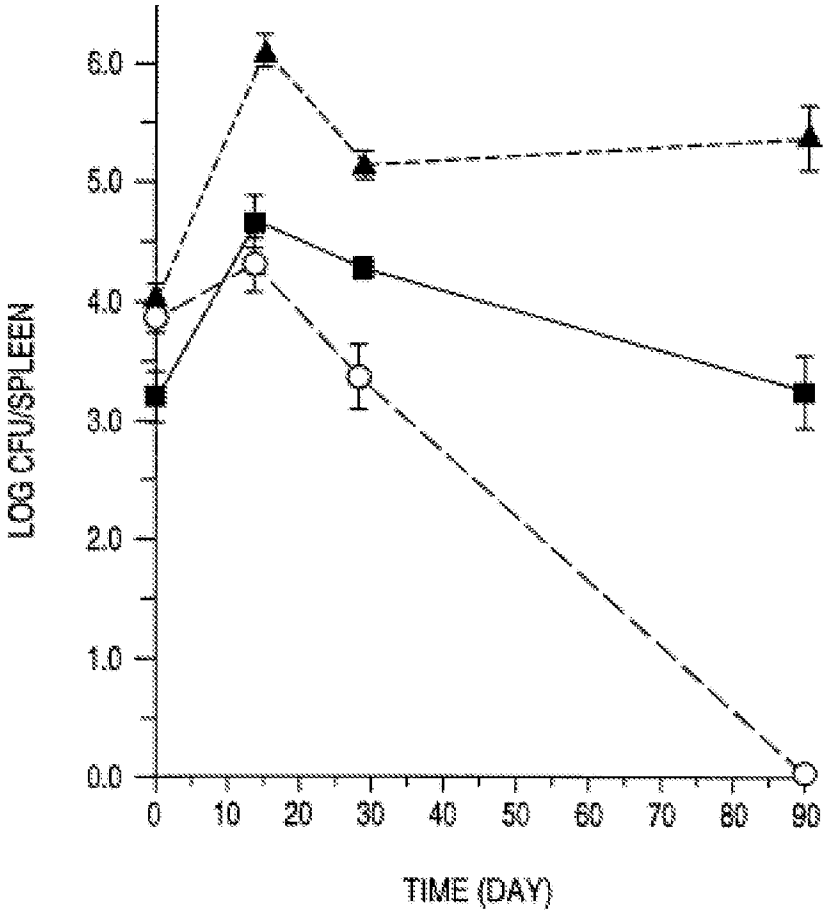
LOG CFU/SPLEEN
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0
10
20
30
40
50
60
70
80
90
TIME (DAY)

FIG. 5A

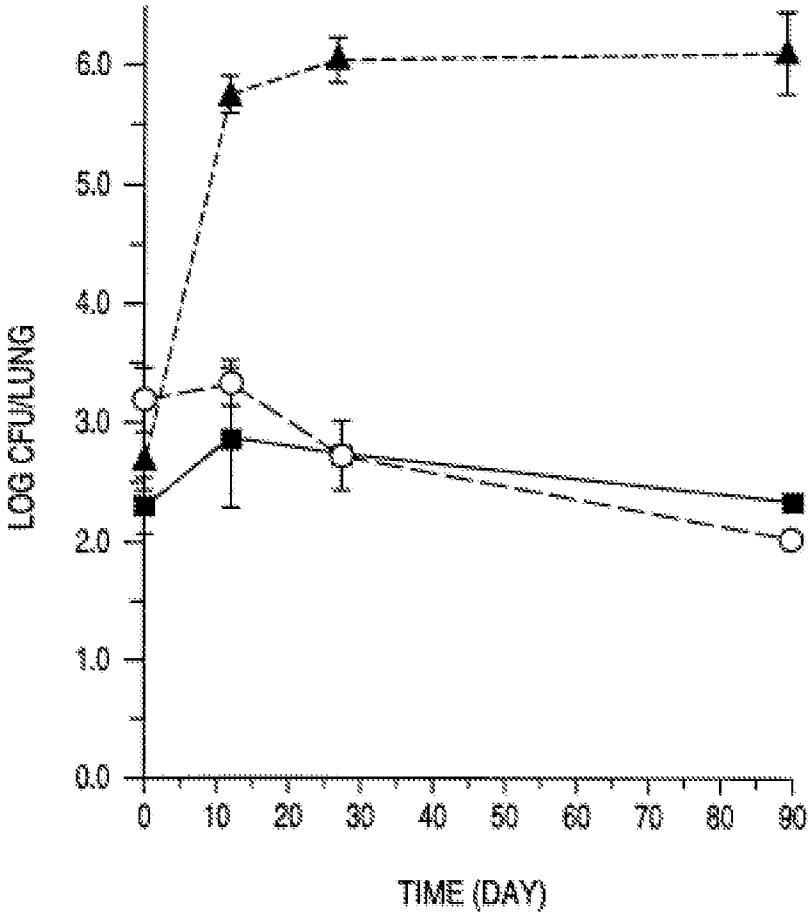
LOG CFU/LUNG
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0
10
20
30
40
50
60
70
80
90
TIME (DAY)

FIG. 5B

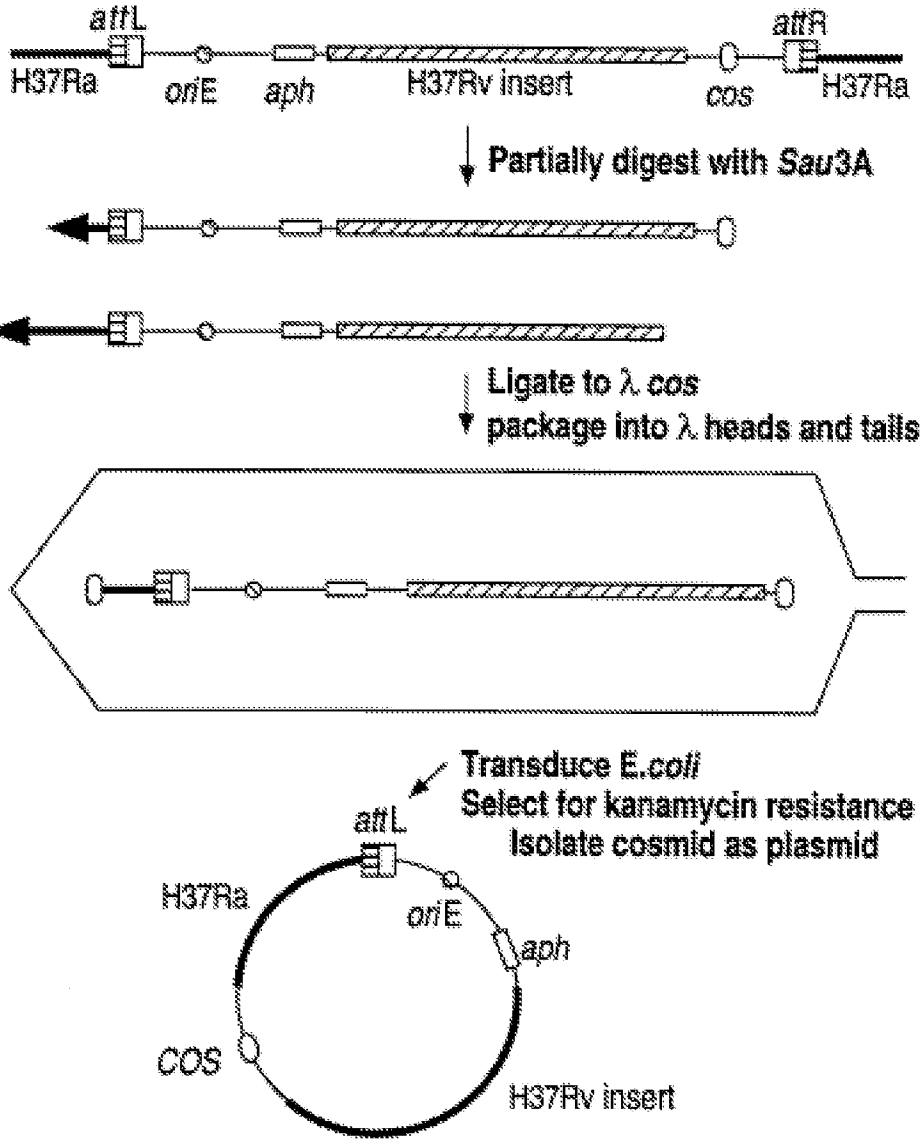
attL
attR
H37Ra
oriE
aph
H37Rv insert
cos
H37Ra
Partially digest with Sau3A
Ligate to λ cos
package into λ heads and tails
Transduce E.coli
Select for kanamycin resistance
Isolate cosmid as plasmid
attL
H37Ra
oriE
aph
COS
H37Rv insert

FIG. 6A

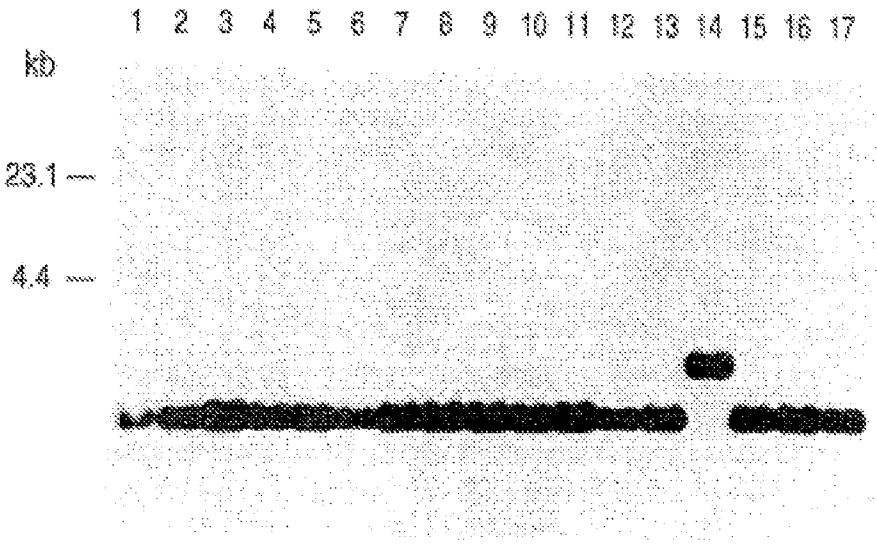

FIG. 6B

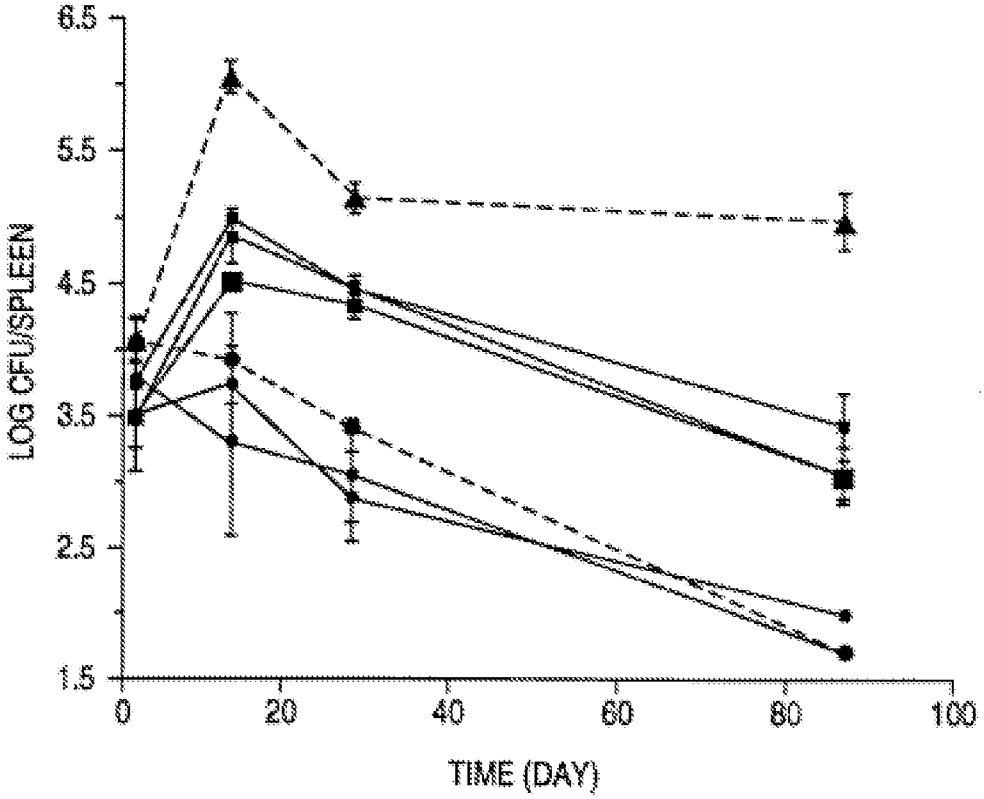

FIG. 7

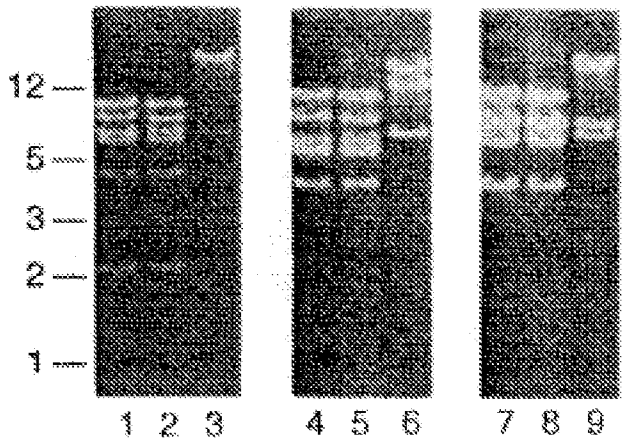

FIG. 8A

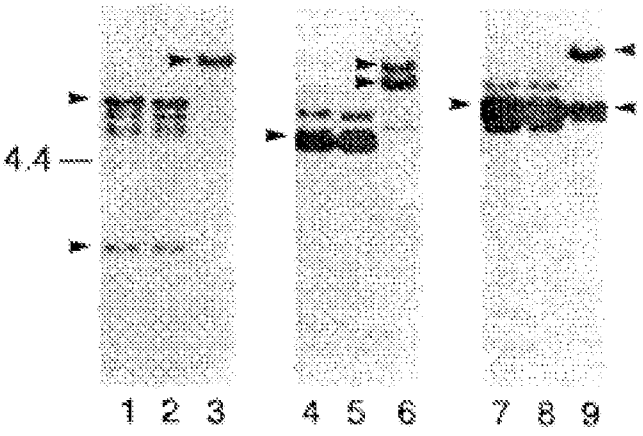

FIG. 8B

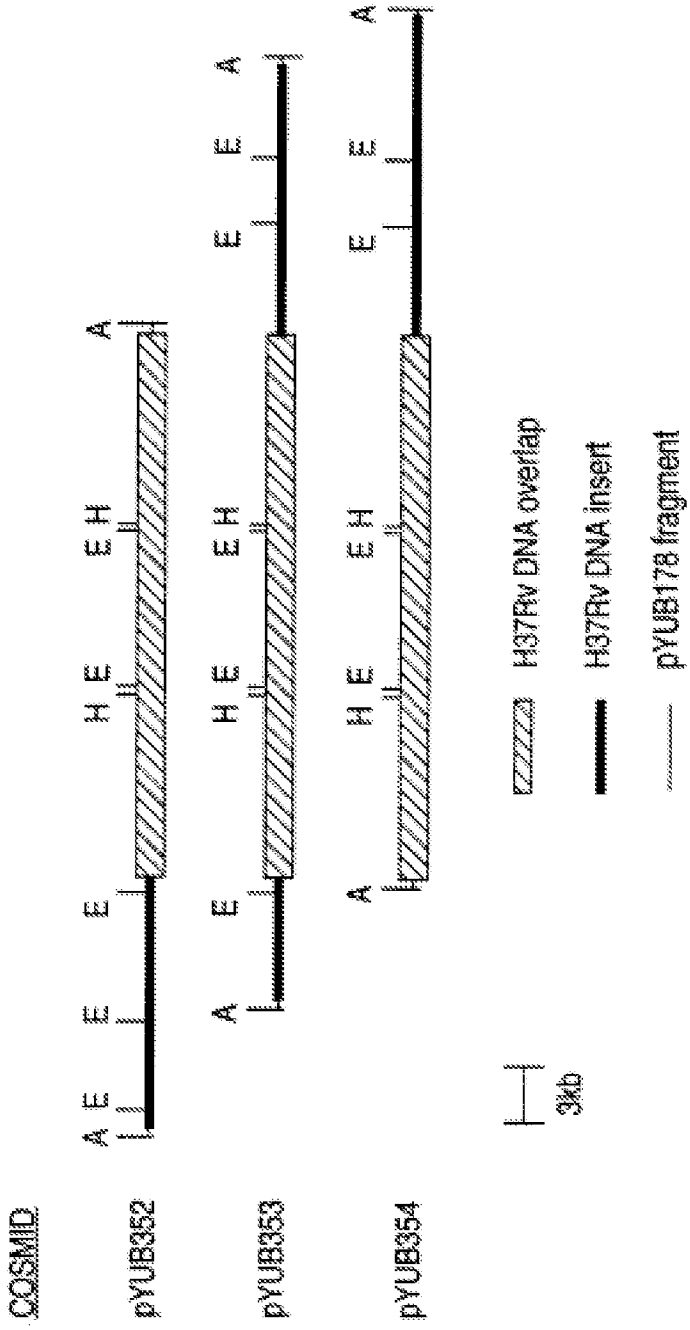
COSMID
pYUB352
pYUB353
pYUB354
A E E E H E H A
A E H E H E E A
A H E E H E E A
3kb
H37Rv DNA overlap
H37Rv DNA insert
pYUB178 fragment

FIG. 8C

```
   1  GATCAAGCTG CTGACCCCGC AACCGGCCAC TCCGTTGGCG GTGCCAAAA
  51  CCATCGCCGA GGTCGTCAAC GGTTTCGGCT GGCGGGGTCC GCTGGGGGTG
 101  ACCTATCCCG GCGTCGTCAC TCACGGCGTC GTCCGGACCG CGGCTAACGT
 151  GGACAAGTCC TGGATAGGGA CCAACGCAGG CGACACTATC GGCGCCGAGC
 201  TGGGCGGTCA GCAGGTCACC ATCCTCAACG ACGCTGATGC CGCCGGCCTG
 251  GCCGAGACAC GCTACGGGGC CGGCAAGAAC AACCCTGGCT TAGTGGTACT
 301  GCTCACATTC GGAACCGGGA TCGGGTCGGC GGTCATCCAC AACGGGACGT
 351  TGATACCCAA CACCGAGTTC GGACATCTTG AGGTCGGCGG CAAGGAAGCG
 401  GAGGAAAGGG CCGCCTCCTC GGTAAAGGAA AAGAACGACT GGACCTATCC
 451  AAAGTCGGCC AAGCAGGTGA CACGCGTGCT CATCGCCATC GAGAACGCGA
 501  TCTCGCCCTGA CCTGTTCATC GCCGGCCGCG GCATCAGCCG CAAGGCCGAC
 551  AAATCGGTGC CGCTACTGGA AAACCGCACA CCAGTAGTGC CCGCGGCCCT
 601  GCAGAACACC GCCGGAATTG TCGGTGCGCC CATCGCCTCT GTGGCAGATA
 651  CGACGCACTG AAACTTGCCC GCTCGGGCTG TACTCGTGCG CAGTAAAGTT
 701  ACAATCGTCA GCGGCGCCCG CCCGACCGAT AGCGCCGAG TATTCACGCT
 751  GATATCAACG CCGACATTCG ACATAGCAGA CACTTTCGGT TACGCACGCC
 801  CAGACCCAAC CGCAAGTGAG TAACGACCGA AGGCGTGTAT GTGGCAGCGA
 851  CCAAAGCAAG CACGGCGACC GATGAGCCCG TAAAACGCAC CGCCACCAAG
 901  TCGCCCCGCG CTTCGGCGTC CGGCGCCAAG ACCGGCCCCA AGCGAACAGC
```

FIGURE 9 - 1

```
951  GGCGAAGTCC GCTAGTGGCT CCCCACCCGC GAAGCGGCCT ACCAAGCCCG
1001 CGGCCCGGTC CGTCAAGCCC GCCTCGGCAC CCCAGGACAC TACGACCAGC
1051 ACCATCCCGA AAAGGAAGAC CCGCGCCGCG GCCAAATCCG CGGCCGCGAA
1101 CGCACCGTCG GCCCGCGCCC ACGCGACCAA GCCACGGCCG CCCAAGGATG
1151 CCCAGCACGA AGCCGCAAGG GATCCCGAGG AGCCCTGGA CTCCGTCGAG
1201 GAGCTCGACG CTGAACCAGA CCTCGACGTC GAGCCCGCCG AGCACCTCGA
1251 CCTTGAGGCC GCCGACCTCA ACCTCGATGA CCTCGAGGAC GAGGTGCGGC
1301 CGGACGCCGA CGACGACCTC GACTCGGCCG ACGACGAAGA CCACGAAGAC
1351 CTCGAAGCTG AGCCGCCGT CGCGCCCGGC CAGACCGCCG ATGACGACGA
1401 GGAGATCGCT GAACCCACCG AAAAGGACAA GGCCTCCGGT GATTCGTCT
1451 GCGATGAAGA CGAGTCGGAG GCCCTCGTC AAGCAGCAA GGACGCCGAA
1501 CTCACCGCAT CCGCCGACTC GGTTCGCGCC TACCTCAAAC AGATCGGCAA
1551 CGTAGCGCTG CTCAACGCCG AGGAAGACGT CGAGCTAGCC CGAGCGATCG
1601 AGCCTGGCCT GTACGCCACG CAGCTGATGA CCGAGCTTAG CGAGCGCGGC
1651 GAAAAGCTGC CTGCCGCCA GCGCCGCGAC ATGATGTGGA TCTGCCGCGA
1701 CGGCGATCGC GCGAAAAACC ATCTGCTGGA AGCCAACCTG CGCCTGGTGG
1751 TTCGCTAGC CAAGCGCTAC ACCGGCCGGG GCATGGCGTT TCTCGACCTG
1801 ATCCAGGAAG GCAACCTGGG GCTGATCCGC GCGGTGGAGA AGTTCGACTA
1851 CACCAAGGGG TACAAGTTCT CCACCTACGC TACGTGGTGG ATTCGCCAGG
```

FIGURE 9 - 2

```
1901  CCATCACCCG CGCCATGGCC GACCAGGCCC GCACCATCCG CATCCCGGTG
1951  CACATGGTCG AGGTGATCAA CAAGCTGGGC CGCATTCAAC GCGAGCTGCT
2001  GCAGGACCTG GGCCGGAGCC CCACGCCCGA GGAGCTGGCC AAAGAGATGG
2051  ACATCACCCC GGAGAAGGTG CTGGAAATCC AGCAATACGC CGGCGAGCCG
2101  ATCTCGTTGG ACCAGACCAT CGGCGACGAG GGCGACAGCC AGCTTGGCGA
2151  TTTCATCGAA GACAGCGAGG CGGTGGTGGC CGTCGACGCG GTGTCCTTCA
2201  CTTTGCTGCA GGATCAACTG CAGTCGGTGC TGGACAGCCT CTCCGAGCGT
2251  GAGGCGGCCG TGGTGCTGCT ACGCTTCGGC CTTACCGACG GCCAGCCGCC
2301  CACCCTTGAC GAGATCGCCC AGGTCTACGG CGTGACCGGG GAACGCATCC
2351  GCCAGATCGA ATCCAAGACT ATGTCGAAGT TGCGCCATCC GAGCCGCTCA
2401  CAGGTCCTGC GCGACTACCT GGACTGAGAG CGCCCGCCGA GGCGACCAAC
2451  GTAGCGCGCC CCCATGTCAG CTAGCCGCAC CATGGTCTCG TCCGGATCGG
2501  AGTTCGAATC AGCCGTCGCC TACTCGCCCG CGGTACCCAT CGGGCCACTC
2551  GTGGTGGTGG CCGCAACGAC CGGCAGCGGC GATGATATCG TCGCTCAGAC
2601  GCGAGACGCT CTGGGCCGCA TCGAGATTGC GCTCGGACAG GCCGGCGCAA
2651  CTCTGGCCGA CGTGGTCCGT ACCCGCATCT ATGTGACCGA TATTCCCGC
2701  TGGCGGAGGG TCGGCGAAGT GCATGCACAG GCTTTCGGCA AGATC
```

FIGURE 9 - 3

```
1    GATCAAGCTGCTGACCCGCAAGCTGGCCACTCGTTGGCGGTCGGCAAAACCATCGCCGA   60
                                                                    A
61   GGTCGTCAACGGTTTCGGCTGCCGCGGTCCGCTGCCCGGTGACCTATCCCCGCGTCGTCAC  120

121  TCACGCGTCGTCCTGCCACCGCTGCTAACGTTGGACAAGTCCTGGATAGCGGACCAACGCACG  180

181  CGACACTATCGGCGCCGAGCTGGGCGGTCAGCAGGTCACCATCCTCAACGACGCTGATGCC  240

241  CGCCCGGCTGGCCGACACGCTACGCGCCCGCAAGAACAACCCTGGCTTAGTGGTACT  300

301  GCTCACATTCGGAACCGGCGATCGGGTCCGCGGTCATCCACAACGGGACGTTGATACCCAA  360

361  CACCGAGTTCGGACATCTTGACGGTCGCCGGCAAGGAAGGGAGGAAAGGGCCGCCTCCTC  420

421  GGTAAAGCAAAAGAACGACTCCCACCTATCCAAAGTGCCCCAAGCAGGTGACACGCGTGCT  480

481  CATCGCCATCGAGAAGGTCGATCTCGCCTGACCTGTTCATCGCTCGGCGCGGCATCAGCCG  540

541  CAAGGCCGACAAATGGGTGCCGCTACTGGAAAACCGCACACCAGTAGTGACCCGCCCCT  600

601  GCAGAACACCGCCCCGAATTGTCGGTGCGCCATGGCCTCTGTCGCAGATACGACGCACTG  660

661  AAACTTGCCCGCTCGGCCTGTACTCGTCCCAGTAAAGTTACAATGGTCAGCGGCGCCG  720

721  CCCGACCGATAGCGCCGAGTATTCACGCTGATATCAACGCCGACATTCGACATAGCCAGA  780

781  CACTTTCCGTTACGCACGCCCAGACCCAACCCGAAGTGAGTAACGACCGAAGGGGTGTAT  840
                                                                    V Y

841  GTCGCAGCCGACCAAAGCAAGCACCGCGACCGATGAGCCGGTAAAACGCACCGCCACCAAG  900
     V A A T K A S T A T D E P V K R T A T K

901  TCGCCCGCGCTTCCGGTCCGGGCCAAGACCGCCCCAAGCGACAGCGCGCGAAGTCC  960
     S P A A S A S G A K T G P K R R T A A K S
```

FIGURE 9A - 1

```
961  GCTAGTGGCTCCCCACCCGCGAAGCGGCCTACCAAGCCCCCCCCCCCCCCCCCGTCCGTCAAGCCC 1020
       A  S  G  S  P  P  A  K  R  A  T  K  P  A  A  R  S  V  K  P

1021 GCCTCCCCCACCCCAGGACACTACGACCAGCACCATCCCGAAAAGGAAGACCCGCGCCGCG 1080
       A  S  A  P  Q  D  T  I  T  I  S  I  I  P  K  R  K  T  R  A  A

1081 GCCAAATCCGCCCCCCCGAAGCACCGTCGCCCCGCGCCAGCCAAGCCACGGGCG 1140
       A  K  S  A  A  A  K  A  P  S  A  R  G  H  A  T  K  P  R  A

1141 CCCAAGGATGCCCAGCACGAAGCCCCAACGGATCCCGAGGACGCCCTGGACTCCGTCGAG 1200
       P  K  D  A  Q  H  E  A  A  T  D  P  E  D  A  L  D  S  V  E

1201 GAGCTCGACGCTGAACCAGACCTCGACGTCGAGCCCGGCGAGGACCTCGACCTTGACGCC 1260
       E  L  D  A  E  P  D  L  D  V  E  P  G  E  D  L  D  L  D  A

1261 GCCGACCTCAACCTCGATGACCTCGACGACGACGTGGCGCCGGACGCCGACGACGACCTC 1320
       A  D  L  N  L  D  D  L  E  D  D  V  A  P  D  A  D  D  D  L

1321 GACTCGGCGACGACGAAGACACGAAGACCTCGAAGCTGAGCGCGGCCGTCGCGCCCGCC 1380
       D  S  G  D  D  E  D  H  E  D  L  E  A  E  A  A  V  A  P  G

1381 CAGACCGCCCGATGACGACGAGGAGATCGCTGAACCACCGAAAAGGACAAGGCCTCCGGT 1440
       Q  T  A  D  D  D  E  E  I  A  E  P  T  E  K  D  K  A  S  G

1441 GATTTCGTCTGGGATGAAGACGAGTCGGAGCCCTGGTCAAGCACGCAAGGACGCCGAA 1500
       D  F  V  W  D  E  D  E  S  E  A  L  R  Q  A  R  K  D  A  E

1501 CTCACCGCATCCGCCGACTCGGTTCGCGCCTACCTCAACAGATGGCAAGGTAGCGCTG 1560
       L  T  A  S  A  D  S  V  R  A  Y  L  K  Q  I  G  K  V  A  L
```

FIGURE 9A - 2

```
1561  CTCAACGCCGAGCAAGAGGTCGAGCTACCCAAGCTGGATCGAGGCTGGCCTGTACGCCAG 1620
        L  N  A  E  E  E  V  E  L  A  K  K  R  I  E  A  G  L  Y  A  T
1621  CAGCTGATGACCGAGCTTAGCCGAGCTGGCGGAAAAGCTGCCTGCCCAGCGCCGGAC 1680
        Q  L  M  T  E  L  S  E  R  G  E  K  L  P  A  A  Q  R  R  D
1681  ATGATGTCGATCTGCCGCGACTGGCGATCGCCGGAAAAACCATCTGCTGGAAGCCAACCTG 1740
        M  M  W  I  C  R  D  G  D  R  A  K  N  H  L  L  E  A  N  L
1741  CGCCTGGTGGTTTCGCTAGCCAAGCGCTACACCGGCCGGGGCATCGAGCTTCTCGACCTG 1800
        R  L  V  V  S  L  A  K  R  Y  T  G  R  G  M  A  F  L  D  L
1801  ATCCAGGAAGGCAACCTGGGGCTGATCCGCGCGTCGAGAAGTTCGACTACACCAAGGGG 1860
        I  Q  E  G  N  L  G  L  I  R  A  V  E  K  F  D  Y  T  K  G
1861  TACAAGTTCTCCACCTACGCTACGTGGTGGATTCGACCAGGCCATCACCGAGCCATGGCC 1920
        Y  K  F  S  T  Y  A  T  W  W  I  R  Q  A  I  T  R  A  M  A
1921  GACCAGGCCCTGCACCATCCGCATCCGGTGCACATCGTCGAGGTGATCAACAAGCTGGGC 1980
        D  Q  A  R  T  I  R  I  P  V  H  M  V  E  V  I  N  K  L  G
1981  CGCATTCAAGCGGAGCTGCTGCAGGACCTGGGCCGCGAGCCCACGCCCGAGGAGCTGGCC 2040
        R  I  Q  R  E  L  L  Q  D  L  G  R  E  P  T  P  E  E  L  A
2041  AAAGAGATGGACATCACCCCGAGAAGGTGCTGGAAATCCAGCAATACGCCGCGCGAGCCG 2100
        K  E  M  D  I  T  P  E  K  V  L  E  I  Q  Q  Y  A  R  E  P
2101  ATCTCGTTGGACCAGACCATCGGCGACGAGGAGGACAGCCAGCTTGGCGATTTCATCGAA 2160
        I  S  L  D  Q  T  I  G  D  E  G  D  S  Q  L  G  D  F  I  E
```

FIGURE 9A - 3

```
2161  GACAGCGAGCCGGTGGTGGCCGTCGAGGCGGTGTCCTTCACTTTGCTGCAGGATCAACTG  2220
        D  S  E  A  V  V  A  V  D  A  V  S  F  T  L  L  Q  D  Q  L
2221  CAGTCGTTGCTGGACACGCTCTCCGAGCGTGAGCCGCCGCGTCGTGCGGCTACGCTTCGGC  2280
        Q  S  V  L  D  T  L  S  E  R  E  A  G  V  V  R  L  R  F  G
2281  CTTACCGACGGCCAGCCGCGCACCCTTGACGAGATCGGCCAGGTCTACGCCGTGACCCGG  2340
        L  T  D  G  Q  P  R  T  L  D  E  I  G  Q  V  Y  G  V  T  R
2341  GAACGCATCCGCCAGATCGAATCCAAGACTATGTCGAAGTTGCGCATCCGAGCCGCTCA  2400
        E  R  I  R  Q  I  E  S  K  T  M  S  K  L  R  H  P  S  R  S
2401  CAGGTCCTGCTGCGACTACCTGGACTGAGAGCGCCCGCCGACCGACCAAGGTAGCCGGCC  2460
        Q  V  L  R  D  Y  L  D  *
2461  CCCATGTCACCTAGCCGCACCATGGTCTCGTCCGGATCGGAGTTCGAATCAGCCGTCGGC  2520
2521  TACTCGCGCGCCGGTAGCGATCGGCCACTCGTGGTGGTGGCCGCGAACGACCGGCAGCCGC  2580
2581  GATGATATCGTCGCTCAGACGCGAGACGCTCTGGCCGCATCGAGATTGCGCTCGGACAG  2640
2641  GCCGGCGCAACTCTGGCCGAGGTCGTCGTACCCGCATCTATGTGACCGATATTTCCCGC  2700
2701  TGGCGCGAGGTCGGCGAAGTGCATGCACAGGCTTTCGGCAAGATC                2745
```

FIGURE 9A - 4

```
                1                                                  50
M.bovis rpoV    VYVAA..... TKASTATDEP VKRTATKSPA ASASGAKTGP KRTAAKSASG
S.coelicolor hrdB MVSAAE.PKR TRKSVAAKSP AKRTATKAVA ANPVTSRKA. .....TAP
S.griseus hrdB  MVSAAESPKR ARKSVAAKSP VKRTATKTVA AKTTVTRTV. .....AAT

               51                                                 100
M.bovis rpoV    SPPAKRATKP AARSVKPASA PQDTTSTIP KRKTRAAAKS AAMKAPSARS
S.coelicolor hrdB AAPAAPATEP AAVE.EEAPA K..... .....KA AAKKTTAKKA
S.griseus hrdB  AAPAVESADA ADDAVAAAPA K..... .....KT AAKKATAKKA

              101                                                 150
M.bovis rpoV    HATKPRAPKD AQNEAATDPE DALDSVEELD AEPDLDVEPG EDLDLDMADL
S.coelicolor hrdB TAKKTTAKKA AAKKTTAKKE DGELLEDEAT EEPKA..ATE EPEGTENAGF
S.griseus hrdB  AAKKTTAKKT AAKK.SGKDD DELLDGDEAA EEVKAGKGEE EEGEGEMKGF

              151                                                 200
M.bovis rpoV    NLLDLEDDVA PDADDDLDSG DDEDHEDLEA EAAVAPGQTA DDDEELAEPT
S.coelicolor hrdB VLSDEDEDDA P..... ..... ..... ..... ..... .....
S.griseus hrdB  VLSDDDEDDA P..... ..... ..... ..... ..... .....
```

FIGURE 10A - 1

```
                201                                                 250
M.bovis rpoV    EKDKASGDFV WDEDESEALR QARKDAELTA SADSVRAYLK QIGKVALLNA
S.coelicolor hrdB ..........  .......... .AQDVAAAGA TADPVKDYLK QIGKVPLLNA
S.griseus hrdB  ..........  .......... .AQDVAVAGA TADPVKDYLK QIGKVPLLNA

                251                                                 300
M.bovis rpoV    EEEVELAKRI EAGLYATQLM TELSERGEKL PAAQKRDMVW ICRDGIRAKN
S.coelicolor hrdB EQEVELAKRI EAGLFAEDKL AN....SDKL APKLKRELEI IAEDGRRAKN
S.griseus hrdB  EQEVELAKRI EAGLFAEDKL AN....ADKL APKLKRELEI IAEDGRRAKN

                301                                                 350
M.bovis rpoV    HLLEAMLRLV VSLAKRYTGR GMAFLDLIQE GNLGLIRAVE KFDYTKGYKF
S.coelicolor hrdB HLLEAMLRLV VSLAKRYTGR GMLFLDLIQE GNLGLIRAVE KFDYTKGYKF
S.griseus hrdB  HLLEAMLRLV VSLAKRYTGR GMLFLDLIQE GNLGLIRAVE KFDYTKGYKF

                351                                                 400
M.bovis rpoV    STYATWWIRQ AITRAMADQA RTIRIPVHMV EVINKLGRIQ RELLQDLGRE
S.coelicolor hrdB STYATWWIRQ AITRAMADQA RTIRIPVHMV EVINKLARVQ RQMLQDLGRE
S.griseus hrdB  STYATWWIRQ AITRAMADQA RTIRIPVHMV EVINKLARVQ RQMLQDLGRE
```

FIGURE 10A - 2

```
                401                                                450
M.bovis rpoV    PTPEELAKEM DITPEKVLEI QQYAREPISL DQTIGDEGDS QLGDFIEDSE
S.coelicolor hrdB PTPEELAKEL DMTPEKVIEV QKYGREPISL HTPLGEDGDS EFGDLIEDSE
S.griseus hrdB  PTPEELAKEL DMTPEKVIEV QKYGREPISL HTPLGEDGDS EFGDLIEDSE

                451                                                500
M.bovis rpoV    AWAVDAVSF TLLQDQLQSV LDTLSEREAG VWRLRFGLTD GQPKTLDEIG
S.coelicolor hrdB AVVPADAVSF TLLQEQLHSV LDTLSEREAG VVSMRFGLTD GQPKTLDEIG
S.griseus hrdB  AVVPADAVSF TLLQEQLHSV LDTLSEREAG VVSMRFGLTD GQPKTLDEIG

                501                                536
M.bovis rpoV    QVYGVTRERI RQIESKTMSK LRHPSRSQVL RDYLD*
S.coelicolor hrdB KVYGVTRERI RQIESKTMSK LRHPSRSQVL RDYLD*
S.griseus hrdB  KVYGVTRERI RQIESKTMSK LRHPSRSQVL RDYLD*
```

FIGURE 10A - 3

```
   Gap Weight:  3.000      Average Match:  0.540
Length Weight:  0.100   Average Mismatch: -0.396

      Quality:  262.3            Length:    536
        Ratio:  0.699             Gaps:       8
Percent Similarity: 72.632   Percent Identity: 59.649

108071.pep x cont.pepf    May 30, 1994  12:52  ..

  1 MVSAAESPKRARKSVAAKSPVKRTATKTVA.....AKTTVTRTVA..... 40
    :  ||. :     |.|...||||||||..|     |||...||.|
  1 VYVAATXA.....STATDEPVKRTATKSPAASASGAKTGPKRTAAKSASG 45

 41 .........ATAAPAVESADAADDAVAAAPAK...KTAAKKATAKKAAAKK 79
                 ||..|..|.|::|..... :|   :.|||.|.||.:|:
 46 SPPAKRATKPAARSVKPASAPQDTTTSTIPKRKTRAAAKSAAAKAPSARG 95

 80 TTAKKTAAKK.......................................... 89
      :.|. |:|.
 96 HATKPRAPKDAQHEAATDPEDALDSVEELDAEPDLDFEPGEDLDLDAADL 145

 90 ....................SGKQDDEILQGDEAAEEVKAGKGEEEEGEGE 120
                         ::.:|.| |:::.|.....:.::|| :|..
146 NLDDLEDDVAPDADDDLDSGDDEDHEDLEAEAAVAPGQTADDDEEIAEPT 195

121 NK....GFVLSDDDEDDA..PAQQVAVAGATADPVKDYLKQIGKVPLLNA 164
    :|    |.:.|:||.:|  .|..  |  .|.||.|:|||||||:|||
196 EKDKASGDFVWDEDESEALRQARKDAELTASADSVRAYLKQIGKVALLNA 245

165 EQEVELAKRIEAGLFAEDKLAN....ADKLAPKLKRELEIIAEDGRRAKN 210
    |:|||||||||||:|.: :::     ::||::  :|::  |. || ||||
246 EEEVELAKRIEAGLYATQLMTELSERGEKLPAAQRRDMMWICRDGDRAKN 295

211 HLLEANLRLVVSLAKRYTGRGMLFLDLIQEGNLGLIRAVEKFDYTKGYKF 260
    ||||||||||||||||||||||| ||||||||||||||||||||||||||
296 HLLEANLRLVVSLAKRYTGRGMAFLDLIQEGNLGLIRAVEKFDYTKGYKF 345

261 STYATWWIRQAITRAMADQARTIRIPVHMVEVINKLARVQRQMLQDLGRE 310
    ||||||||||||||||||||||||||||||||
346 STYATWWIRQAITRAMADQARTIRIPVHMV..................... 375
```

FIGURE 11

```
b                    1                                                  50
M.bovis ATCC35721    VYVAA.....  TKASTATDEP VKRTATKSPA ASASGAKTGa KRTAAKSASG
M.bovis WAg200,WAg201 VYVAA.....  TKASTATDEP VKRTATKSPA ASASGAKTGp KRTAAKSASG
M.tuberculosis Erdman VYVAA.....  TKASTATDEP VKRTATKSPA ASASGAKTGa KRTAAKSASG
S.coelicolor         mvsAAe.pkr  TrkSvAakap aKRTATKavA Anpvtsrka. ........tap
S.griseus            mvsAAespkr  arkSvAakap VKRTATKtvA Aktcvtrtv. ........Aat

                     51                                                 100
M.bovis ATCC35721    SPPAKRATKP  AARSVKPASA PQDTTTSTIP KRKTRAAAKS AAAKAPSARG
M.bovis WAg200,WAg201 SPPAKRATKP  AARSVKPASA PQDTTTSTIP KRKTRAAAKS AAAKAPSARG
M.tuberculosis Erdman SPPAKRATKP  AARSVKPASA PQDTTTSTIP KRKTRAAAKS AAAKAPSARG
S.coelicolor         aaPAapATEP  AAve.eeApA k.......... ..........Ka AAkKttakka
S.griseus            aaPAvesada  AddaVaaApA k.......... ..........Kt AAkKAtakka

                     101                                                150
M.bovis ATCC35721    HATKPRAPKD  AQHEAATDPE DALDSVEELD AEPDLDVEPG EDLDLDAADL
M.bovis WAg200,WAg201 HATKPRAPKD  AQHEAATDPE DALDSVEELD AEPDLDVEPG EDLDLDAADL
M.tuberculosis Erdman HATKPRAPKD  AQHEAATDPE DALDSVEELD AEPDLDVEPG EDLDLDAADL
S.coelicolor         tAkKttAkKa  Aakkttakke Dgelledeat eEPka..ate EpegtemAgf
S.griseus            aAkKttAkKt  Aakk.egkgd Delldgdeaa eEvkagkgee Eegegenkgf
```

FIGURE 12 - 1

```
                    151                                                200
M.bovis ATCC35721   NLDDLEDDVA PDADDDLDSG DDEDHEDLEA EAAVAPGQTA DDDEEIAEPT
M.bovis WAg200,WAg201 NLDDLEDDVA PDADDDLDSG DDEDHEDLEA EAAVAPGQTA DDDEEIAEPT
M.tuberculosis Erdman NLDDLEIDVA PDADDDLDSG DDEDHEDLEA EAAVAPGQTA DDDEEIAEPT
S.coelicolor        vLsDdeDdA P.......... .......... .......... ..........
S.griseus           vLsDdteDdA P.......... .......... .......... ..........

                    201                                                250
M.bovis ATCC35721    EKDKASGDFV WDEDESEALR QARKDAELTA SADSVRAYLK QIGKVALLNA
M.bovis WAg200,WAg201 EKDKASGDFV WDEDESEALR QARKDAELTA SADSVRAYLK QIGKVALLNA
M.tuberculosis Erdman EKDKASGDFV WDEDESEALR QARKDAELTA SADSVRAYLK QIGKVALLNA
S.coelicolor        .......... .......... .AqqvAaagA tADpVkdYLK QIGKVpLLNA
S.griseus           .......... .......... .AqqvAvagA tADpVkdYLK QIGKVpLLNA

                    251                                                300
M.bovis ATCC35721   EEEVELAKRI EAGLYATQLM TELSERGEKL PAAQRRDMMW ICRDGDRAKN
M.bovis WAg200,WAg201 EEEVELAKRI EAGLYATQLM TELSERGEKL PAAQRRDMMW ICRDGDRAKN
M.tuberculosis Erdman EEEVELAKRI EAGLYATQLM TELSERGEKL PAAQRRDMMW ICRDGDRAKN
S.coelicolor        EqEVELAKRI EAGLfAedkl am.....sdKL apklkReIei IaeDGrRAKN
S.griseus           EqEVELAKRI EAGLfAedkl am.....adKL apklkReIei IaeDGrRAKN
```

FIGURE 12 - 2

```
                    301                                                 350
M.bovis ATCC35721   HLLEANLRLV VSLAKRYTGR GMAFLDLIQE GNLGLIRAVE KFDYTKGYKF
M.bovis WAg200,WAg201 HLLEANLRLV VSLAKRYTGR GMAFLDLIQE GNLGLIRAVE KFDYTKGYKF
M.tuberculosis Erdman HLLEANLRLV VSLAKRYTGR GMAFLDLIQE GNLGLIRAVE KFDYTKGYKF
S.coelicolor        HLLEANLRLV VSLAKRYTGR GMIFLDLIQE GNLGLIRAVE KFDYTKGYKF
S.griseus           HLLEANLRLV VSLAKRYTGR GMIFLDLIQE GNLGLIRAVE KFDYTKGYKF

                    351                                                 400
M.bovis ATCC35721   STYATMWIRQ AITRAMADQA RTIRIPVHMV EVINKLGRIQ RELLQDLGRE
M.bovis WAg200,WAg201 STYATMWIRQ AITRAMADQA RTIRIPVHMV EVINKLGRIQ RELLQDLGRE
M.tuberculosis Erdman STYATMWIRQ AITRAMADQA RTIRIPVHMV EVINKLGRIQ RELLQDLGRE
S.coelicolor        STYATMWIRQ AITRAMADQA RTIRIPVHMV EVINKLaRvQ RqmLQDLGRE
S.griseus           STYATMWIRQ AITRAMADQA RTIRIPVHMV EVINKLaRvQ RqmLQDLGRE

                    401                                                 450
M.bovis ATCC35721   PTPEELAKEM DITPEKVLEI QQYAREPISL EQTIGDEGDS QLGDFIEDSE
M.bovis WAg200,WAg201 PTPEELAKEM DITPEKVLEI QQYAREPISL EQTIGDEGDS QLGDFIEDSE
M.tuberculosis Erdman PTPEELAKEM DITPEKVLEI QQYAREPISL EQTIGDEGDS QLGDFIEDSE
S.coelicolor        PTPEELAKEI DmTPEKViEv QkYgREPISL htplGedGDS efGDlIEDSE
S.griseus           PTPEELAKEI DmTPEKViEv QkYgREPISL htplGedGDS efGDlIEDSE
```

FIGURE 12 - 3

```
                    451                                                500
M.bovis ATCC35721   AVVAVDAVSF TLLQDQLQSV LDTLSEREAG VVRLRFGLTD GQPRTLDEIG
M.bovis WAg200,WAg201 AVVAVDAVSF TLLQDQLQSV LDTLSEREAG VVRLRFGLTD GQPRTLDEIG
M.tuberculosis Erdman AVVAVDAVSF TLLQDQLQSV LDTLSEREAG VVRLRFGLTD GQPRTLDEIG
S.coelicolor        AVVpaDAVSF TLLQeQLhSV LDTLSEREAG VVsmRFGLTD GQPkTLDEIG
S.griseus           AVVpaDAVSF TLLQeQLhSV LDTLSEREAG VVsmRFGLTD GQPkTLDEIG

                    501                 ↓          536
M.bovis ATCC35721   QVYGVTRERI RQIESKTMSK LrHPSRSQWL RDYLD*
M.bovis WAg200,WAg201 QVYGVTRERI RQIESKTMSK LRHPSRSQWL RDYLD*
M.tuberculosis Erdman QVYGVTRERI RQIESKTMSK LRHPSRSQWL RDYLD*
S.coelicolor        kVYGVTRERI RQIESKTMSK LRHPSRSQWL RDYLD*
S.griseus           kVYGVTRERI RQIESKTMSK LRHPSRSQWL RDYLD*
```

FIGURE 12 - 4

a

GATCAAGCTGCTGACCCGCAACCAGCCACTCCGTTGGCGGTCGCCAAAACCATCGCCGA 60

GGTCGTCAACGGTTTCGGCTGGCGAGGGTCCGCTGGGGATGACCTATCCGGGCGTCGTCAC 120

TCACGGCGTCGTCCGACCGCCGCTAACGTGGACAAGTCCTGGATAGGGACCAACGCACG 180

CGACACTATCGGCAGCCGAGCTGGGCGGTCAGCAGGTCACATCCTCAAGGACGCTGATGC 240

CGCCGGGCTGGCCGACACCCTACGGGAGCCAAGAACAACCCTGGCTTAGTGGTACT 300

GCTCACATTCGGAACCGGGATCGGGTCCGGTCATCCACAACGGGACGTTGATACCCAA 360

CACCGAGTTCGGACATCTTGAGGTCGGCGGCAAGCAAGCGGAAAGCTGCCCTCCTC 420

CGTAAAGGAAAAGAACGACTGGACCTATCCAAAGTGGCCAAGCAGGTGACACGGTGCT 480

CATCGCCATCGAGAACGGATCTGGCCTGACCTGTTCATCGCCGGCGCGCTGCATCAGCCG 540

CAAGGCCGACAAATGCGTGCCGCTACTGGAAAACCGCACACCAGTAGTGCCGCGGCCCT 600

GCAGAACACCGCCGGAATTGTCGGTGCGCCATGGCCTCTGTTCGCAGATACGACGCACTG 660

AAACTTGCCCGCTCCGGCTGTACTCGTGCGCAGTAAAGTTACAATCGTCAGCGGCGGCCG 720

CCCGACCGATAGCGGCGAGTATTCACGCTGATATCAACGCCGACATTCGACATAGCAGA 780

CACTTTCGGTTACGCACGCCCAGACCCAAGCCGGAAGTGAGTAACGACCGAAGGCGTGTAT 840

GTGGCAGGCGACCAAAGCAAGCACGGCGACCGATGAGCCGGTAAAACGCACCGCCAGCAAG 900

V A A I K A S T A T D E P V K R T A T K

FIGURE 12a - 1

G (35721 and Erdman)

```
TCGCCCAGCGCCTTCCGGCGTCCCGGCCAAGACCGGCCCCAAGCGAACAGCGGCGAAGTCC   960
  S  P  A  A  S  A  S  G  A  K  T  G  P  K  R  T  A  A  K  S
                                 A
GCTAGTGGCTCCCCAGCCGCGAAGGCGCCTACCAAGCCCGCGGCCGGTCCGTCAAGCCC  1020
  A  S  G  S  P  P  A  K  R  A  T  K  P  A  A  R  S  V  K  P
GCCTCGCCACCCCAGGACACTACGACCAGCACCATCCCGAAAGCGAAGACGGCGCCGCG  1080
  A  S  A  P  Q  D  T  T  T  S  T  I  P  K  R  K  T  R  A  A
GCCAAATCCGCCCCCCCCGAAGCCACCGTCCCCCCCCCGCCACGCGACCAAGCCACGGGCG  1140
  A  K  S  A  A  A  K  A  P  S  A  R  G  H  A  T  K  P  R  A
CCCAAGCATGCCCAGCACGAAGCCGCAACCGATCCCGAGGACGCCCTGGACTCCGTCGAG  1200
  P  K  D  A  Q  H  E  A  A  T  D  P  E  D  A  L  D  S  V  E
GAGCTCGAAGCTGAACCAGACCTCGACGTCGACCCCGGCGAGGACCTCGACCTTGACGCC  1260
  E  L  D  A  E  P  D  L  D  L  D  V  E  P  G  E  D  L  D  L  D  A
GCCGACCTCAACCTCGATGACCTCGAGGAGGACGTGGGCGCCGACGCCGACGACGACCTC  1320
  A  D  L  N  L  D  D  L  E  D  D  V  A  P  D  A  D  D  D  L
GACTCCGGCGACGACGAAGACCAGGAAGACCTCGAAGCTGAGGCCGCCGTCGGCGCCGGC  1380
  D  S  G  D  D  E  D  H  E  D  L  E  A  E  A  A  V  A  P  G
CAGACCGCCGATGACGAGGAGATCGCTGAACCCACCGAAAAGGACAAGGCCTCCGGT  1440
  Q  T  A  D  D  E  E  I  A  E  P  T  E  K  D  K  A  S  G
```

FIGURE 12a - 2

```
GATTTCGTCTGGGATGAAGACGAGTCGGAGGCCCTGCGTCAAGCACGCAAGGACGCCGAA 1500
 D  F  V  W  D  E  D  E  S  E  A  L  R  Q  A  A  R  K  D  A  E
CTCACCGCATCCGCCGACTCGGTTCGGCGCTACCTCAAACAGATGGCAAGGTAGCGCTG 1560
 L  I  A  S  A  D  S  V  R  A  A  Y  L  K  Q  I  G  K  V  A  L
CTCAAGCGCGAGGAAGAGGTCGAGCTAGCCAAGCGGATCGAGCTGCCCTGTACGCCACG 1620
 L  N  A  E  E  E  V  E  L  A  K  R  I  E  A  G  L  Y  A  T
CAGCTGATGACCGAGCTTAGCCAGGGCGGCGAAAAGCTGCCTGCCGCCCAGCGCCGCGAC 1680
 Q  L  M  T  E  L  S  E  R  G  E  K  L  P  A  A  Q  R  R  D
ATGATGTCGATCTGCCGCGACGGCGACGCGATCGCCGAAAAGCCATCTGCTGGAAGCCAACCTG 1740
 M  M  W  I  C  R  D  G  D  R  A  K  N  H  L  L  E  A  N  L
CGCCTGGTGGTTTCGCTAGCCAAGCGCTACAGCGGCGGCATGGAGTTTCTCGACCTG 1800
 R  L  V  V  S  L  A  K  R  Y  T  G  R  G  M  A  F  L  D  L
ATCCAGGAAGGCAACCTGGGGCTGATCCGCGGTGGAGAAGTTCGACTACACCAAGGGG 1860
 I  Q  E  G  N  L  G  L  I  R  A  V  E  K  F  D  Y  T  K  G
TACAAGTTCTCCACCTACGCTACGTGGTGGATTCGCCAGGCCATCACCCGCGCCATGGCC 1920
 Y  K  F  S  T  Y  A  T  W  W  I  R  Q  A  I  T  R  A  M  A
GACCAGGCCCGCACCATCCGCATCCCGGTGCACATGGTCGAGGTGATCAACAAGCTGCGC 1980
 D  Q  A  R  T  I  R  I  P  V  H  M  V  E  V  I  N  K  L  G
CGCATTCAACGCGAGCTGCTGCAGGACCTGGCCCGCGAGCCCACGCCCGAGGAGCTGGCC 2040
 R  I  Q  R  E  L  L  Q  D  L  G  R  E  P  T  P  E  E  L  A
```

FIGURE 12a - 3

```
AAAGAGATGGACATCACCCCGGAGAGGTGCTGGAAATCCAGCAATACGCCCGCGAGCCG 2100
 K  E  M  D  I  T  P  E  K  V  L  E  I  Q  Q  Y  A  R  E  P
ATCTGTTGGACCAGACCATCGGCGACGAGGGCGACAGCCAGCTTGGCGATTTCATCGAA 2160
 I  S  L  D  Q  T  I  G  D  E  G  D  S  Q  L  G  D  F  I  E
GACAGCGAGGGCGGTGGTGGCGTCGACGCGGTGTCCTTCACTTTGCTGCAGGATCAACTG 2220
 D  S  E  A  V  V  A  V  D  A  V  S  F  T  L  L  Q  D  Q  L
CAGTCGGTGCTGCACACGCTCTCCGAGCGTGAGGCGGGCGTGGTGGCGCTACGCTTCGGC 2280
 Q  S  V  L  D  T  L  S  E  R  E  A  G  V  V  R  L  R  F  G
CTTACCGACGGCCAGCCGCGCACCCTTGACGAGATCGGCCAGGTCTACGGCGTGACCCGG 2340
 L  T  D  G  Q  P  R  T  L  D  E  I  G  Q  V  Y  G  V  T  R
                    A (35721)
GAACGCATCCGCCAGATCGAATCCAAGACTATGTCGAAGTTGCGCCATCCGAGCCGCTCA 2400
 E  R  I  R  Q  I  E  S  K  T  M  S  K  L  R  H  P  S  R  S
                          H
CAGGTCCTGGAGGACTACCTGGACTGAGAGGCGCCCGCCGAGGCGACCAACGTAGCGGGCC 2460
 Q  V  L  R  D  Y  L  D  *
CCCATGTCAGCTAGCCGCACCATGGTCTCGTCCGCATCGGAGTTCGAATCAGCCGTCGGC 2520
TACTCGCGCGCGGTACGCATCGGCCACTCGTGGTGGTCGGAACGACCGGCAGCGGC 2580
          C (WAg201, 35721 and Erdman)
GATGATATCGTCGCTCAGACGCGAGACGCCTCTGCGCCGCATCGAGATTGCGCTCGGACAG 2640
```

FIGURE 12a - 4

GCCGCGCAACTCTGGCCGACGTCGTACCCGCATCTATGTGACCGATATTTCCCGC 2700
TGGCGCGAGGTCGGCGAAGTGCATGCACAGGCTTTCGGCAAGATC 2745

FIGURE 12a - 5